Package 'MEB'

April 15, 2020

Type Package

Title A Minimum Enclosing Ball (MEB) method to detect differential expression genes for RNA-seq data

Version 1.0.0

Author Yan Zhou, Jiadi Zhu

Maintainer Jiadi Zhu <2160090406@email.szu.edu.cn>, Yan Zhou <zhouy1016@szu.edu.cn>

Description Identifying differential expression genes between the same or different species is an urgent demand for biological research. In most of cases, normalization is the first step to solve this problem, then by employing the hypothesis testing, we could detect statistically significant genes. With the development of machine learning, it gives us a new perspective on discrimination between differential expression (DE) and non-differential expression (non-DE) genes. Provided a set of training data, the procedure of distinguishing genes could be simplified as a classification problem. However, in reality, it is hard for us to get the information from both DE and non-DE genes. To solve this problem, we try to identify differential cases only in the domain of non-DE genes, and transform the problem to an outlier detection in machine learning.
Given that non-DE genes have some similarities in features, we build a Minimum Enclosing Ball (MEB) to cover those non-DE genes in feature space, then those DE genes, which are enormously different from non-DE genes, being regarded as outliers and rejected outside the ball. Compared with existing methods, it is no need for the MEB method to normalize data in advance. Besides, the MEB method could be easily applied to the same or different species data and without changing too much.

License GPL-2

Encoding UTF-8

LazyData true

Depends R (>= 3.6.0)

Suggests knitr,rmarkdown

VignetteBuilder knitr

RoxygenNote 6.1.1

Imports e1071, SummarizedExperiment

biocViews DifferentialExpression, GeneExpression, Normalization, Classification

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

git_branch RELEASE_3_10

git_last_commit 8d8a47b

git_last_commit_date 2019-10-29

Date/Publication 2020-04-14

R topics documented:

`MEB` *Detect differential expression genes for RNA-seq data*

Description

Use the Minimum Enclosing Ball (MEB) method to discriminate differential expression (DE) genes in the same or different species.

Usage

```
MEB(countsTable, train_id, gamma, nu = 0.01, reject_rate = 0.1,
ds = FALSE)
```

Arguments

`countsTable`	Matrix or data.frame of short read counts for each genes in the same or different species.
`train_id`	A vector shows the position of housekeeping genes or conserved genes in countsTable.
`gamma`	A parameter needed for all kernels except linear.
`nu`	parameter needed for one-classification.
`reject_rate`	A value used in controling the scale of ball, default is 0.01.
`ds`	A value to show the data is for the same species or different species. If ds is FALSE, the data is the same species, else the data is the different species.

Value

list(.) A list of results, "model" represents the model of MEB, which could be used to discriminate a new gene, "gamma" represents the selected gamma parameters in model MEB, "train_error" represents the corresponding train_error when the value of gamma changed.

Examples

```
## Simulation data for the same species.
library(SummarizedExperiment)
data(sim_data_sp)
gamma <- seq(1e-06,5e-05,1e-06)
sim_model_sp <- MEB(countsTable = assay(sim_data_sp), train_id=1:1000,
gamma, nu = 0.01, reject_rate = 0.05, ds = FALSE)

## Real data for the same species.
data(real_data_sp)
gamma <- seq(1e-06,5e-05,1e-06)
real_model_sp <- MEB(countsTable = assay(real_data_sp), train_id=1:530,
gamma, nu = 0.01, reject_rate = 0.1, ds = FALSE)

## Simulation data for the different species.
data(sim_data_dsp)
gamma <- seq(1e-07,2e-05,1e-06)
sim_model_dsp <- MEB(countsTable = assay(sim_data_dsp), train_id=1:1000,
gamma, nu = 0.01, reject_rate = 0.1, ds = TRUE)

## Real data for the different species.
data(real_data_dsp)
gamma <- seq(5e-08,5e-07,1e-08)
real_model_dsp <- MEB(countsTable = assay(real_data_dsp), train_id=1:143,
gamma, nu = 0.01, reject_rate = 0.1, ds = TRUE)
```

`real_data_dsp` *A real dataset of genes between the different species.*

Description

This data set includes two species and 19330 genes with corresponding short read counts, in which the first 143 genes are conserved genes.

Usage

```
real_data_dsp
```

Format

A data.frame contains two species and 19330 genes.

Source

Brawand, D., Soumillon, M., Necsulea, A. and Julien, P. et al. (2011). The evolution of gene expression levels in mammalian organs. Nature, 478, 343-348.

real_data_sp *A real dataset of genes between the same species.*

Description

This data set includes two samples and each sample with five biological replicates and 16519 genes with corresponding short read counts, in which the first 530 genes are housekeeping genes.

Usage

```
real_data_sp
```

Format

A data.frame contains two samples and each sample with five biological replicates and 16519 genes.

Source

Marioni J.C., Mason C.E., et al. (2008). RNA-seq: an assessment of technical reproducibility and comparisonwith gene expression arrays. Genome Res. 18(9), 1509–1517.

sim_data_dsp *A simulation dataset of genes between the differetn species.*

Description

This data set includes two species and 18472 genes with corresponding short read counts, in which the first 1000 genes are conserved genes.

Usage

```
sim_data_dsp
```

Format

A data.frame contains two species and 18472 genes.

Source

Jiadi Zhu, Yan Zhou, Junhui Wang, Youlong Yang (2019, pending publication). A minimum enclosing ball method to detect differential expression genes for RNA-seq data.

`sim_data_sp` *A simulation dataset of genes between the same species.*

Description

This data set includes two samples and 10623 genes with corresponding short read counts, in which the first 1000 genes are housekeeping genes.

Usage

```
sim_data_sp
```

Format

A data.frame contains two samples and 10623 genes.

Source

Jiadi Zhu, Yan Zhou, Junhui Wang, Youlong Yang (2019, pending publication). A minimum enclosing ball method to detect differential expression genes for RNA-seq data.

Index