Package ‘Cormotif’

November 20, 2016

Type Package

Title Correlation Motif Fit

Version 1.20.0

Date 2011-06-08

Author Hongkai Ji, Yingying Wei

Maintainer Yingying Wei <ywei@jhsph.edu>

Description It fits correlation motif model to multiple studies to
detect study specific differential expression patterns.

Depends R (>= 2.12.0), affy, limma

Imports affy, graphics, grDevices

License GPL-2

LazyLoad yes

biocViews Microarray, DifferentialExpression

NeedsCompilation no

R topics documented:

cormotif internal .................................................. 2
cormotiffit .......................................................... 2
cormotiffitall ....................................................... 4
cormotiffitfull ..................................................... 5
cormotiffitsep ...................................................... 6
generank ............................................................ 7
plotIC ................................................................. 8
plotMotif ........................................................... 9
simudata2 ........................................................... 10

Index 11
Correlation Motif Internal functions

Description

These functions are not part of the package application programming interface and are not recommended to be used by the users.

Usage

modt.f0.loglike
modt.f1.loglike
cmfit
cmfitall
cmfitsep
cmfitfull
limmafit
generatetype

References


cormotifit

Correlation Motif Fit

Description

This function fits the Correlation Motif model to multiple expression studies. It gives the fitted values for the probability distribution of each motif, the fitted values of the given correlation matrix and the posterior probability for each gene to be differentially expressed in each study.

Usage

cormotifit(exprs, groupid, compid, K=1, tol=1e-3, max.iter=100, BIC=TRUE)

Arguments

exprs a matrix, the expression data after normalization that is on log2 scale, each row of the matrix corresponds to a gene and each column of the matrix corresponds to a sample array.
groupid the group label for each sample array, two arrays in the same study with same experiment condition(e.g.control) have the same groupid.
compid the study design and comparison matrix, each row of the matrix corresponds to one study with the first column being the first experiment condition and the second column being the second experiment condition.
K a vector, each element specifying the number of motifs a model wants to fit.
tol the relative tolerance level of error.
Details

For the $i^{th}$ element of $K$, the function fits total number of $K[i]$ motifs to the data. Each gene can belong to one of the $K[i]$ possible motifs according to prior probability distribution, $\text{motif.prior}$. For genes in motif $j$, the probability that they are differentially expressed in study $d$ is $\text{motif.q}(j, d)$. One should indicate the groupid and compid for each study clearly.

Value

- `bestmotif$p.post` the posterior probability for each gene to be differentially expressed in each study for the best fitted model
- `bestmotif$motif.prior` fitted values of the probability distribution of different motifs for the best fitted model
- `bestmotif$motif.q` fitted values of the correlation motif matrix for the best fitted model
- `bestmotif$loglike` log-likelihood of the best fitted model
- `bic` the BIC values of all fitted models
- `aic` the AIC values of all fitted models
- `loglike` log-likelihood of all fitted models

Author(s)

Hongkai Ji, Yingying Wei

References


Examples

data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
#the expression data is from the second column to m
exprs.simu2<-as.matrix(simudata2[,2:m])
#prepare the group label for each sample array
data(simu2_groupid)
#prepare the design matrix for each group of samples
data(simu2_compgroup)
#fit 2 correlation motifs to the data
motif.fitted<-cormotiffit(exprs.simu2, simu2_groupid,simu2_compgroup,K=2)
cormotiffitall

All Studies Correlation Motif Fit

Description

This function assumes that a gene is either differentially expressed in all studies or is not differentially expressed in any study. It gives the fitted values for the probability distribution of motif (0,0,...,0) and motif (1,1,...,1), and the posterior probability for each gene to be differentially expressed in all studies.

Usage

cormotiffitall(exprs, groupid, compid, tol=1e-3, max.iter=100)

Arguments

- **exprs**: a matrix, the expression data after normalization that is on log2 scale, each row of the matrix corresponds to a gene and each column of the matrix corresponds to a sample array.
- **groupid**: the group label for each sample array, two arrays in the same study with same experiment condition (e.g. control) have the same groupid.
- **compid**: the study design and comparison matrix, each row of the matrix corresponds to one study with the first column being the first experiment condition and the second column being the second experiment condition.
- **tol**: the relative tolerance level of error.
- **max.iter**: maximum number of iterations.

Details

The difference between `cormotiffitall` and `cormotif(..., K = 2, ...)` is that `cormotiffitall` forces the motif to be one of the two patterns but `cormotif` allows motif patterns other than (0,...,0) and (1,...,1).

Value

- **p.post**: the posterior probability for each gene to be differentially expressed
- **motif.prior**: fitted values of the probability distribution of motif (0,0,...,0) and motif (1,1,...,1)
- **loglike**: log-likelihood of the fitted model

Author(s)

Hongkai Ji, Yingying Wei

References

Examples

data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
  #the expression data is from the second column to m
exprs.simu2<-as.matrix(simudata2[,2:m])
  #prepare the group label for each sample array
data(simu2_groupid)
  #prepare the design matrix for each group of samples
data(simu2_compgroup)
  #fit the two motifs (0,0,...0) and (1,1,...,1) to the data
motif.fitted.all<-cormotiffitall(exprs.simu2, simu2_groupid, simu2_compgroup)

cormotiffitfull

Full Model Motif Fit

Description

This function fits the data to the model with all $2^D$ possible 0-1 patterns, where $D$ is the number of studies.

Usage

cormotiffitfull(exprs, groupid, compid, tol=1e-3, max.iter=100)

Arguments

exprs  a matrix, the expression data after normalization that is on log2 scale, each row of the matrix corresponds to a gene and each column of the matrix corresponds to a sample array.

groupid  the group label for each sample array, two arrays in the same study with same experiment condition (e.g. control) have the same groupid.

compid  the study design and comparison matrix, each row of the matrix corresponds to one study with the first column being the first experiment condition and the second column being the second experiment condition

 tol  the relative tolerance level of error.

max.iter  maximum number of iterations.

Details

The difference between cormotiffitfull and cormotif(...) $K = 2^D, ...$ is that cormotiffitfull forces motif to be one of the those 0-1 patterns. For cormotiffit, the motif does not necessarily to be of either 1 or 0, such as (0,1,...,0). It could be (0.9,0.4,...,0.2).

Value

p.post  the posterior probability for each gene to be differentially expressed.

motif.prior  fitted values of the probability distribution of the $2^D$ 0-1 motifs.

loglike  log-likelihood of the fitted model.
Author(s)

Hongkai Ji, Yingying Wei

References


Examples

data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
# the expression data is from the second column to m
exprs.simu2<-as.matrix(simudata2[,2:m])

# prepare the group ID number for each sample array
data(simu2_groupid)

# prepare the design matrix for each group of samples
data(simu2_compgroup)

# fit 2^D 0-1 motifs to the data
motif.fitted.sep<cormotiffitfull(exprs.simu2, simu2_groupid, simu2_compgroup)

---

cormotiffitsep Individual Study Motif Fit

Description

This function fits a mixture modified t-distribution model to each study separately.

Usage

cormotiffitsep(exprs, groupid, compid, tol=1e-3, max.iter=100)

Arguments

exprs a matrix, the expression data after normalization that is on log2 scale, each row of the matrix corresponds to a gene and each column of the matrix corresponds to a sample array.

groupid the group label for each sample array, two arrays in the same study with same experiment condition (e.g. control) have the same groupid.

compid the study design and comparison matrix, each row of the matrix corresponds to one study with the first column being the first experiment condition and the second column being the second experiment condition

tol the relative tolerance level of error.

max.iter maximum number of iterations.
**generank**

**Value**

- `p.post`: the posterior probability for each gene to be differentially expressed.
- `motif.prior`: fitted values of the probability for genes to be differentially expressed in each study, a $1 \times D$ vector, where $D$ is the number of studies.
- `loglike`: log-likelihood of the fitted model.

**Author(s)**

Hongkai Ji, Yingying Wei

**References**


**Examples**

```r
data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
#the expression data is from the second column to m
exprs.simu2<-as.matrix(simudata2[,2:m])

#prepare the group ID number for each sample array
data(simu2_groupid)

#prepare the design matrix for each group of samples
data(simu2_compgroup)

#fit seprate models to each study
motif.fitted.sep<-cormotiffitsep(exprs.simu2, simu2_groupid,simu2_compgroup)
```

---

**generank**: Rank genes based on statistics

**Description**

This function ranks the genes according to the decreasing order of the given statistics.

**Usage**

`generank(x)`

**Arguments**

- `x`: A $G \times D$ matrix of statistics, the number of rows is the number of genes and the number of columns is the number of studies.

**Details**

The function returns a $G \times D$ matrix of index of top ranked genes in each study according to the decreasing order of statistics in that study.
Examples

data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
#the expression data is from the second column to m
exprs.simu2<-as.matrix(simudata2[,2:m])

#prepare the group ID number for each sample array
data(simu2_groupid)

#prepare the design matrix for each group of samples
data(simu2_compgroup)

#fit 2 correlation motif to the data
motif.fitted<-cormotiffit(exprs.simu2, simu2_groupid,simu2_compgroup,K=2)
#give the gene index list according to the decreasing order of
#posterior probability for a gene to be differentially expressed in each study
generank(motif.fitted$bestmotif$p.post)

plotIC

### BIC and AIC plot

Description

This function plots BIC and AIC values for all fitted motif models.

Usage

plotIC(fitted_cormotif)

Arguments

fitted_cormotif

The object obtained from cormotiffit.

Details

The left graph is the BIC plot and the right graph is the AIC plot.

Author(s)

Hongkai Ji, Yingying Wei

References

Examples

```r
data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
# the expression data is from the second column to \text{eqn}(m)
exprs.simu2<-as.matrix(simudata2[,2:m])

# prepare the group ID number for each sample array
data(simu2_groupid)

# prepare the design matrix for each group of samples
data(simu2_compgroup)

# fit 2 correlation motif to the data
motif.fitted<-cormotiffit(exprs.simu2, simu2_groupid,simu2_compgroup,K=2)
plotIC(motif.fitted)
```

---

**plotMotif**  
*Correlation Motif plot*

**Description**

This function plots the Correlation Motif patterns and the associated prior probability distributions.

**Usage**

```
plotMotif(fitted_cormotif,title="")
```

**Arguments**

- `fitted_cormotif`  
The object obtained from `cormotiffit`.
- `title`  
The title for the graph.

**Details**

Each row in both graphs corresponds to one motif pattern. The left graph shows the correlation motif pattern. The grey color scale of cell \((k, d)\) indicates the probability that motif \(k\) is differentially expressed in study \(d\). Each row of the bar chart corresponds to the motif pattern in the same row of the left pattern graph. The length of the bar in the bar chart shows the number of genes of the given pattern in the dataset, which is equal to \(\text{motif.fitted}\$\text{bestmotif}\$\text{motif.prior}\) multiplying the number of total genes.

**Author(s)**

Hongkai Ji, Yingying Wei

**References**

Examples

```r
data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
#the expression data is from the second column to m
exprs.simu2<-as.matrix(simudata2[,2:m])

#prepare the group ID number for each sample array
data(simu2_groupid)

#prepare the design matrix for each group of samples
data(simu2_compgroup)

#fit 2 correlation motif to the data
motif.fitted<-cormotiffit(exprs.simu2, simu2_groupid,simu2_compgroup,K=2)

plotMotif(motif.fitted)
```

---

**simudata2**  
Example dataset for Cormotif

---

**Description**

Here we present three files needed for the various Correlation Motif fit functions.

**Details**

simudata2 are combined from four studies sharing the same 3,000 genes, each having two experiment conditions and three samples for each condition. simudata2 saves the expression values for all genes and all sample arrays on log2 scale; simu2_groupid prepares the group label for each sample; and simu2_compgroup describes the study design.

**References**

Index

*Topic correlation motif fit
  cormotiffit, 2
  cormotiffitfull, 5

  cmfit (cormotif internal), 2
  cmfitall (cormotif internal), 2
  cmfitfull (cormotif internal), 2
  cmfitsep (cormotif internal), 2
  cormotif internal, 2
  cormotiffit, 2
  cormotiffitall, 4
  cormotiffitfull, 5
  cormotiffitsep, 6

  generank, 7
  generatetype (cormotif internal), 2

  limmafit (cormotif internal), 2

  modt.f0.loglike (cormotif internal), 2
  modt.f1.loglike (cormotif internal), 2

  plotIC, 8
  plotMotif, 9

  simu2_compgroup (simudata2), 10
  simu2_groupid (simudata2), 10
  simudata2, 10