

Cormotif

March 24, 2012

`cormotif internal` *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

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

`cormotiffit` *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

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

Arguments

<code>exprs</code>	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.
<code>groupid</code>	the group label for each sample array, two arrays in the same study with same experinment condition(<i>e.g.control</i>) have the same groupid.
<code>compid</code>	the study design and comparison matrix, each row of the matrix corresponds to one study with the first column being the first experinment condition and the second column being the second experinment condition.
<code>K</code>	a vector, each element specifying the number of motifs a model wants to fit.
<code>tol</code>	the relative tolerance level of error.
<code>max.iter</code>	maximun number of iterations.
<code>BIC</code>	default is BIC=TRUE, selecting the model with the lowest BIC value among all fitted models; if BIC=FALSE, selecting the model with the lowest AIC value among all fitted models.

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, *motif.prior*. For genes in motif j , the probability that they are differentially expressed in study d is *motif.q(j, d)*. One should indicate the groupid and compid for each study clearly.

Value

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

Author(s)

Hongkai Ji, Yingying Wei

References

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

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 experinment condition(<i>e.g.control</i>) 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 experinment condition and the second column being the second experinment condition
tol	the relative tolerance level of error.
max.iter	maximun 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 *cormotiffit* 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

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

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 experinment 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 experinment condition and the second column being the second experinment condition

tol the relative tolerance level of error.
 max.iter maximum number of iterations.

Details

The difference between *cormotifitfull* and *cormotif(..., K = 2^D, ...)* is that *cormotifitfull* forces motif to be one of the those 0-1 patterns. For *cormotifit*, 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

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

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<-cormotifitfull(exprs.simu2, simu2_groupid, simu2_compgroup)
```

cormotifitsep *Individual Study Motif Fit*

Description

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

Usage

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

Arguments

<code>exprs</code>	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.
<code>groupid</code>	the group label for each sample array, two arrays in the same study with same experinment condition(<i>e.g.control</i>) have the same groupid.
<code>compid</code>	the study design and comparison matrix, each row of the matrix corresponds to one study with the first column being the first experinment condition and the second column being the second experinment condition
<code>tol</code>	the relative tolerance level of error.
<code>max.iter</code>	maximun number of iterations.

Value

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

Author(s)

Hongkai Ji, Yingying Wei

References

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

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 separte models to each study
motif.fitted.sep<-cormotifitsep(exprs.simu2, simu2_groupid,simu2_compgroup)
```

generank	<i>Rank genes based on statistics</i>
----------	---------------------------------------

Description

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

Usage

```
generank(x)
```

Arguments

`x` A $G * D$ 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 $G * D$ matrix of index of top ranked genes in each study according to the decreasing order of statistics in that study.

Author(s)

Hongkai Ji, Yingying Wei

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 cormotifit.

Details

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

Author(s)

Hongkai Ji, Yingying Wei

References

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

Examples

```
data(simudata2)
n<-nrow(simudata2)
m<-ncol(simudata2)
#the expression data is from the second column to \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<-cormotifit(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 cormotifit.  
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 $motif.fitted$bestmotif$motif.prior$ multiplying the number of total genes.

Author(s)

Hongkai Ji, Yingying Wei

References

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

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<-cormotifit(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

Ji, H., Wei, Y.,(2011) Correlation Motif. Unpublished

Index

*Topic **correlation motif fit**

`cormotiffit`, 1
`cormotiffitfull`, 4

`cmfit` (*cormotif internal*), 1
`cmfitall` (*cormotif internal*), 1
`cmfitfull` (*cormotif internal*), 1
`cmfitsep` (*cormotif internal*), 1
`cormotif internal`, 1
`cormotiffit`, 1
`cormotiffitall`, 3
`cormotiffitfull`, 4
`cormotiffitsep`, 5

`generank`, 7
`generatetype` (*cormotif internal*),
1

`limmafit` (*cormotif internal*), 1

`modt.f0.loglike` (*cormotif
internal*), 1
`modt.f1.loglike` (*cormotif
internal*), 1

`plotIC`, 8
`plotMotif`, 9

`simu2_compgroup` (*simudata2*), 10
`simu2_groupid` (*simudata2*), 10
`simudata2`, 10