Package 'GeneClusterNet'

January 13, 2017

Type Package

Title Gene Expression Clustering and Gene Network

Version 1.0.1

Date 2017-01-11

Depends R(>= 3.1.0)

Imports MASS,G1DBN,igraph

Author Yaqun Wang, Zhengyang Shi and Xiang Zhan

Maintainer Yaqun Wang <yw505@sph.rutgers.edu>

Description Functions for clustering time-course gene expression and reconstructing of gene regulatory network based on Dynamic Bayesian Network.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2017-01-13 17:16:25

R topics documented:

GeneClusterNet-package

cluster gene expressions and reconstruct gene regulatory networks

Description

GeneClusterNet is a contributed R package for reconstructing gene regulatory network from time course gene expression data based on clustering of dynamic gene expressions. It provides functions for gene expression clustering, deciding the optimal number of clusters based on Bayesian Information Criterion (BIC), interpolating expression data for unevenly spaced measurements to have expression data as measured at even time intervals, and applying Dynamic Bayesian Network model to reconstruct gene regulatory networks. It also includes functions for displaying and visualizing clusters and networks.

Details

Package:	GeneClusterNet
Type:	Package
Title:	Gene Expression Clustering and Gene Network
Version:	1.0.1
Date:	2017-01-11
Depends:	R(>= 3.1.0)
Imports:	MASS,G1DBN,igraph
Author:	Yaqun Wang, Zhengyang Shi and Xiang Zhan
Maintainer:	Yaqun Wang <yw505@sph.rutgers.edu>
Description:	Functions for clustering time-course gene expression and reconstructing of gene regulatory network bas
License:	GPL-3
NeedsCompilation:	no

Index of help topics:

Author(s)

Yaqun Wang, Zhengyang Shi and Xiang Zhan

Maintainer: Yaqun Wang <yw505@sph.rutgers.edu>

References

Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., et al. (2012). *How to cluster gene expression dynamics in response to environmental signals*. Briefings in bioinformatics, 13(2), 162-174.

Wang, Y., Berceli, S. A., Garbey, M. and Wu, R. (2016). *Inference of gene regulatory network through adaptive dynamic Beyesian networm modeling*. Technical Report.

R package G1DBN available at https://cran.r-project.org/package=G1DBN

`GeneCluster` *Clustering dynamic gene expression*

Description

Providing functional clustering of time course gene expressions by using the Legendre orthogonal polynomials (LOP) to model cluster-specific curves for each cluster

Usage

```
GeneCluster(mExpression, times, NumberOfCluster, orderLOP)
```

Arguments

`mExpression`	a gene expression matrix with p columns (length of time vector) and n rows (number of genes).
`times`	time vector specifies the time points of measurements.
`NumberOfCluster`	number of cluster (J)
`orderLOP`	order of Legendre Polynomials (r)

Value

A list that contains $MeanExpression is the matrix with J rows (J is number of Cluster) and n columns (length of time vector), each rows is the mean expression of a cluster. $LOPCoefficient is the coefficient matrix of LOP with J rows and r+1 columns.$Classifications indicates the cluster label for each of genes. A list of Plots, first plot is the mean expression plot for every clusters and each of the rest plots displays the trajectories of gene expressions in each cluster.

Author(s)

Yaqun Wang <yw505@sph.rutgers.edu>, Zhengyang Shi

References

Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., et al. (2012). *How to cluster gene expression dynamics in response to environmental signals*. Briefings in bioinformatics, 13(2), 162-174.

Wang, Y., Berceli, S. A., Garbey, M. and Wu, R. (2016). *Inference of gene regulatory network through adaptive dynamic Beyesian networm modeling*. Technical Report.

Examples

```
# load the package
library(GeneClusterNet)

# Set the number of Cluster is 3 and order of Legendre Polynomials is 5.
set.seed(1234)
data(mExpression)

Sample=mExpression[sample(1:nrow(mExpression),50,replace=FALSE),]

GeneCluster(Sample, times=c(1:18), NumberOfCluster=3,orderLOP=5)
```

`GeneClusterBIC` *Optimal number of Gene Clusters*

Description

Given time cousre expressions of n genes, time vector, order of Legendre Polynomials and a range of cluster numbers, e.g. from 1 to 15, the function can identify the optimal number of clusters, which has the smallest BIC value.

Usage

```
GeneClusterBIC(mExpression, times, G = c(1:15), orderLOP)
```

Arguments

`mExpression`	a gene expression matrix with p columns (length of time vector) and n rows (number of genes).
`times`	time vector specifies the time points of measurements.
`G`	range of number of clusters
`orderLOP`	order of Legendre Polynomials

Value

A list of BIC corresponds to every number of clusters and the optimal BIC. A plot shows the smallest BIC.

Author(s)

Yaqun Wang <yw505@sph.rutgers.edu>, Zhengyang Shi

References

Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., et al. (2012). *How to cluster gene expression dynamics in response to environmental signals*. Briefings in bioinformatics, 13(2), 162-174.

Wang, Y., Berceli, S. A., Garbey, M. and Wu, R. (2016). *Inference of gene regulatory network through adaptive dynamic Beyesian networm modeling*. Technical Report.

Examples

```
# load the package
library(GeneClusterNet)
set.seed(1234)
data(mExpression)
Sample=mExpression[sample(1:nrow(mExpression),50,replace=FALSE),]
GeneClusterBIC(Sample, times=c(1:18), G=c(1:5), orderLOP=5)
```

GeneClusterInterp *Interpolating gene expression measurements*

Description

To perform the dynamic Bayesian network analysis, the time points of gene expression measurements have to be evenly spaced. If the original time points in a dataset are not even, this function can not only interpolate measurements to have evenly spaced time course gene expressions, but also allow users to specify the number of time points.

Usage

```
GeneClusterInterp(LOPCoefficient, OriginalTime, outLen = 20)
```

Arguments

`LOPCoefficient`	coefficient matrix of LOP.
`OriginalTime`	time vector
`outLen`	number of new time points.

Value

A matrix with J + 1 rows (J is number of Cluster) and outLen columns (number of equal time space). The first row of it provides the new time vector.

Author(s)

Yaqun Wang <yw505@sph.rutgers.edu>, Zhengyang Shi

References

Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., et al. (2012). *How to cluster gene expression dynamics in response to environmental signals*. Briefings in bioinformatics, 13(2), 162-174.

Wang, Y., Berceli, S. A., Garbey, M. and Wu, R. (2016). *Inference of gene regulatory network through adaptive dynamic Beyesian networm modeling*. Technical Report.

Examples

```
# load the package
library(GeneClusterNet)
set.seed(1234)
data(mExpression)
Sample=mExpression[sample(1:nrow(mExpression),50,replace=FALSE),]
LOPCoefficient =GeneCluster (Sample, times=c(1:18), NumberOfCluster=3,orderLOP=5)$ LOPCoefficient
GeneClusterInterp (LOPCoefficient, OriginalTime=c(1:18),outLen=20)
```

`GeneClusterNet` *Inference of gene regulatory network*

Description

This function clusters gene expressions and makes inference of gene regulatory network with dynamic Bayesian network method based on G1DBN. It integrates the analyses of GeneCluster, GeneClusterBIC, GeneClusterInterp and G1DBN.

Usage

```
GeneClusterNet(mExpression, times, orderLOP, alpha1 = 0.5, alpha2 = 0.05,
realign = F, cutoff = c(lowCut = -0.35, upCut = 0.2),
NumberOfCluster = 0, sLabels = NULL)
```

Arguments

`mExpression`	a gene expression matrix with p columns (length of time vector) and n rows (number of genes).
`times`	time vector
`orderLOP`	order of Legendre Polynomials
`alpha1`	threshold that use for edge selection in the 1st order dependence score matrix S1 and its default setting is 0.5. See https://cran.r-project.org/web/packages/G1DBN/G1DBN.pdf.

alpha2	threshold that use for edge selection in the score matrix S2 and its default setting is 0.05. See https://cran.r-project.org/web/packages/G1DBN/G1DBN.pdf
realign	defalut is FALSE, realign is true, the cutoff variable indicates a cutoff range
cutoff	cutoff range for determining the time of initial up or down regulaiotn.
NumberOfCluster	number of cluster. if it is set to zero, the function will determine the optimal number of gene expression clusters automatically.
sLabels	defalut is NULL, the user could specify the cluster labels when number of cluster is specified.

Value

A score matrix and an adjacency matrix. It also creates a list of plots same as the function GeneCluster. In addtition, it creates a plot of gene regulatory network.

Author(s)

Yaqun Wang <yw505@sph.rutgers.edu>, Zhengyang Shi

References

Wang, Y., Xu, M., Wang, Z., Tao, M., Zhu, J., Wang, L., et al. (2012). *How to cluster gene expression dynamics in response to environmental signals*. Briefings in bioinformatics, 13(2), 162-174.

Wang, Y., Berceli, S. A., Garbey, M. and Wu, R. (2016). *Inference of gene regulatory network through adaptive dynamic Beyesian networm modeling*. Technical Report.

R package G1DBN available at https://cran.r-project.org/package=G1DBN

Examples

```
# load the package
library(GeneClusterNet)
set.seed(1234)
data(mExpression)
Sample=mExpression[sample(1:nrow(mExpression),50,replace=FALSE),]
GeneClusterNet (Sample, times=c(1:18), orderLOP=5 ,alpha1=0.5, alpha2=0.05,NumberOfCluster = 3)
```

---

mExpression *Gene expression data set of Yeast.*

Description

Sample data with 789 genes and 18 evenly spaced gene expression measurements.

Usage

```
data("mExpression")
```

References

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., ... & Futcher, B. (1998). *Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization*. Molecular biology of the cell, 9(12), 3273-3297.

Examples

```
# load the package
library(GeneClusterNet)

data(mExpression)

# first 5 rows

Sample=mExpression[1:5,]

plot(1:18,Sample[1,],type="l", ylim=c(min(Sample),max(Sample)),xlab="Time",ylab="Expression",lwd=2)
color=1
for (i in 2:5){
  color=color+1
  lines(1:18,Sample[i,],col=color,lwd=2)
}
```

Index