Package 'TSNet'

March 18, 2018

Type Package

Title A New Method for Constructing Tumor Specific Gene Co-Expression
Networks Based on Samples with Tumor Purity Heterogeneity

Version 0.1.0

Date 2018-1-1

Depends R (>= 3.0.0)

Imports MASS, mvtnorm, clusterGeneration, doMC, foreach, glasso,
igraph

Description Estimation of tumor purity from mixed expression profiles and estimation of co-
expression networks for tumor and non-tumor components.

License GPL (>= 2)

Maintainer Francesca Petralia <francesca.petralia@mssm.edu>

RoxygenNote 6.0.1

NeedsCompilation no

Author Francesca Petralia [aut, cre],
Pei Wang [aut]

R topics documented:

---

deNet   *Needs one-line description of package.*

Description

Estimation of co-expression networks for tumor and non-tumor components for a certain value of
penalty parameters.

Usage

```
deNet(exprM,purity,paraM,rhoy,rhoz)
```

Arguments

exprM
: A (p x n) matrix of expression with p being the number of genes and n the number of samples.

purity
: An n-dimensional vector containing estimated purity for each sample.

paraM
: A (p x 4) matrix. The first and second columns contain the initial value of the mean of genes in tumor and non-tumor cells, respectively, and the third and fourth columns contain standard deviations for each gene in tumor and non-tumor cells, respectively.

rhoy
: Graphical lasso penalty parameter for covariance matrix in tumor cells.

rhoz
: Graphical lasso penalty parameter for covariance matrix in normal cells.

Value

r
: list containing parameters estimated by the model (i.e., Sigmay, Sigmaz, Uy, Uz)

LL.temp
: log-likelihood of the data for different EM iterations.

BIC
: Bayesian Information Criteria.

LL
: Log-likelihood at convergence.

EY
: Number of edges of tumor network.

EZ
: Number of edges of non-tumor network.

convergence
: value indicating whether or not the algorithm converged

purity
: estimated tumor purity.

data
: expression data matrix.

time
: total running time.

References

Francesca Petralia, Li Wang , Jie Peng , Arthur Yan, Jun Zhu and Pei Wang, A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity (Submitted).

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432-441, 2008.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373-397, 2014.

Examples

```
set.seed(1)
p<-20; # -- number of genes
n=10;  # -- number of samples
rhoy<-rhoz<-.8 # -- penalty parameters
paraM<-cbind(matrix(0,p,2),matrix(1,p,2))
```

```
# --- Sample Data
exprM=matrix(rnorm(p*n),p,n)
purity=runif(n)

# --- Estimate Network
deNet(exprM,purity,paraM,rhoy,rhoz)
```

---

deNet_purity   *Function to estimate tumor purity from mixed expression data.*

Description

TSNet function to estimate tumor purity from mixed expression profiles.

Usage

```
deNet_purity(exprM,purity)
```

Arguments

exprM	A (n x p) matrix of gene expression with p being the number of genes and n the number of samples.
purity	An n-dimensional vector containing tumor-purity for each sample.

Value

Epurity	Estimated tumor purity via TSNet.
paraM	(p x 2) matrix containing mean parameter of tumor and non-tumor components for each gene. First column contains mean of tumor component, while the second column the mean of non-tumor component.
convergence	value indicating whether or not the algorithm converged
time	total running time.

References

Francesca Petralia, Li Wang , Jie Peng , Arthur Yan, Jun Zhu and Pei Wang, A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity (Submitted).

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432-441, 2008.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373-397, 2014.

Examples

```
set.seed(1)
p<-20; n=10;

# --- Sample Data
set.seed(1)
exprM<-matrix(rnorm(p*n),n,p)

meanP<-0.5;
varP<-0.04;
sizeP<-meanP*(1-meanP)/varP-1
realP<-rbeta(n,meanP*sizeP,(1-meanP)*sizeP)

# --- Estimate Purity
deNet_purity(as.matrix(exprM),realP)
```

Evaluate_network | *Performance evaluation in estimating tumor specific and non-tumor specific networks.*

Description

Calculation of true positive and false positive edges in network estimation.

Usage

```
Evaluate_network(TSNet,Data)
```

Arguments

TSNet	Object returned by function deNet.
Data	Object returned by function Synthetic_data.

Value

TPTumor	True Positive Edges estimated for tumor component.
FPTumor	False Positive Edges estimated for tumor component.
TPNormal	True Positive Edges estimated for non-tumor component.
FPNormal	False Positive Edges estimated for non-tumor component.

References

Francesca Petralia, Li Wang , Jie Peng , Arthur Yan, Jun Zhu and Pei Wang, A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity (Submitted).

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432-441, 2008.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373-397, 2014.

Examples

```
set.seed(1)
n=10     # -- number of samples for each replicate
rep=1   # -- number of replicates
p=30    # -- network dimension
betaSize=1

# --- Sample Data
Data<-Synthetic_data(p=p,rep=rep,n=n,cluster=2,
        parameterseed=1,dataseed=1,meanY=NULL,
        meanZ=NULL,sdY=NULL,sdZ=NULL, P.v=NULL,
        meanP=NULL,varP=NULL,sameModuleN=0,betaModel=TRUE,
        topology="Power",betaSize=betaSize)

# --- Estimate Tumor purity
purity<-deNet_purity(exprM=Data$X,purity=Data$P)

# --- Estimate Network
#rhoy<-rhoz<-.8 # -- penalty parameters graphical lasso
#TSNet<-deNet(exprM=t(Data$X),purity=purity$Epurity,
#             paraM=purity$paraM,rhoy=rhoy,rhoz=rhoz)

# --- Evaluate Network
#Evaluate_network(TSNet,Data)
```

Synthetic_data   *Generate independent datasets from a mixed expression model.*

Description

This function is utilized to sample mixed expression profiles.

Usage

```
Synthetic_data(p,rep,n,cluster,parameterseed,dataseed,meanY=NULL,
               meanZ=NULL,sdY=NULL,sdZ=NULL,P.v=NULL,meanP=NULL,
               varP=NULL,sameModuleN=0,topology="Power",
               betaModel=TRUE,betaSize=Inf)
```

Arguments

p	Number of genes (i.e., number of nodes in the network).
rep	Number of independent data sets (i.e., replicates).
n	Number of observations for each replicate.
cluster	Number of disjoint components in the network.
parameterseed	
	Random seed for parameter sampling.
dataseed	Random seed for observations sampling.

meanY	p dimensional vector containing the mean parameter of the Gaussian distribution of the gene expression in tumor component. When NULL, meanY=0.
meanZ	p dimensional vector containing the mean parameter of the Gaussian distribution of the gene expression in non-tumor component. When NULL, meanZ=0.
sdY	p dimensional vector containing the variance parameter of the Gaussian distribution of the gene expression in tumor component. When NULL, sdY=1.
sdZ	p dimensional vector containing the variance parameter of the Gaussian distribution of the gene expression in non-tumor component. When NULL, sdZ=1.
P.v	True level of tumor-purity. When NULL, the value of tumor purity is simulated from a Beta distribution with mean meanP and variance varP.
meanP	Mean of the Beta distribution utilized to simulate tumor purity. When NULL, meanP=0.5.
varP	Variance of the Beta distribution utilized to simulate tumor purity. When NULL, varP=0.04.
sameModuleN	Number of clusters shared between the networks of tumor and non-tumor components.
topology	Which topology between Star and Power should be utilized to simulate tumor-specific and non-tumor specific networks. Default value is Power.
betaModel	TRUE if the observed tumor purity (prior) is drawn from a beta distribution. FALSE if a logistic regression is utilized.
betaSize	Parameter controlling the variance of the distribution of the prior (observed) tumor purity. The mean parameter is equal to the true tumor purity. Default value equal to Inf (i.e., observed tumor purity is equal to the true tumor purity)

Value

List object containing the following values:

X	(p x (n * rep)) matrix of gene expression from mixed model.
Uy	(p x 1) vector containing mean value for each gene in tumor component.
Uz	(p x 1) vector containing mean value for each gene in non-tumor component.
Sigmay	(p x p) matrix containing covarian matrix of tumor component.
Sigmaz	(p x p) matrix containing covarian matrix of non-tumor component.
P	(n x 1) vector containing observed values of tumor purity.
Y	(p x (n * rep)) matrix containing expression profile of tumor component.
Z	(p x (n * rep)) matrix containing expression profile of non-tumor component.
realP	((n * rep) x 1) vector containing true values of tumor purity.

References

Francesca Petralia, Li Wang , Jie Peng , Arthur Yan, Jun Zhu and Pei Wang, A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity (Submitted).

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432-441, 2008.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373-397, 2014.

Examples

```
n=10    # -- number of samples for each replicate
rep=5   # -- number of replicates
p=10    # -- network dimension
betaSize=1

Synthetic_data(p=p,rep=rep,n=n,cluster=2,parameterseed=1,
               dataseed=1,meanY=NULL, meanZ=NULL,sdY=NULL,sdZ=NULL,
               P.v=NULL,meanP=NULL,varP=NULL, sameModuleN=0,
               betaModel=TRUE,topology="Power",betaSize=betaSize)
```

Synthetic_purity	*Performance evaluation in estimating tumor-purity based on synthetic data.*

Description

Generate synthetic data sets and estimate tumor purity for each of the synthetic data sets.

Usage

```
Synthetic_purity(p, Rep, n, cluster, sameModuleN, dataseed,
                 topology, betaModel=TRUE, betaSize=Inf)
```

Arguments

p	Number of genes.
Rep	Number of replicate data sets.
n	Number of observations for each replicate.
cluster	Number of disjoint network components.
sameModuleN	Number of overlapping modules between tumor and non-tumor components.
dataseed	Random seed.
topology	Which topology between "Star" and "Power" should be utilized to simulate tumor specific and non-tumor specific networks.
betaModel	TRUE if the observed tumor purity (prior) is generated from a beta distribution. FALSE if a logistic regression needs to be utilized.
betaSize	Parameter controlling the variance of the distribution of the prior (observed) tumor purity. The mean parameter is equal to the true tumor purity. Default value equal to Inf (i.e., observed tumor purity is equal to the true tumor purity).

Value

List object including:

TSNet.purity	TSNet's estimate of tumor purity. Vector of dimension (Rep * n).
priorP	Prior value of tumor purity. Vector of dimension (Rep * n).
realP	True value of tumor purity. Vector of dimension (Rep * n).

`cor.rep`	Correlation between estimated purity and prior value of tumor purity. Vector of dimension rep.
`cor.true`	Correlation between estimated purity and true value of tumor purity. Vector of dimension rep.

References

Francesca Petralia, Li Wang , Jie Peng , Arthur Yan, Jun Zhu and Pei Wang, A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity (Submitted).

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432-441, 2008.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373-397, 2014.

Examples

```
n=10    # -- number of samples for each replicate
Rep=5   # -- number of replicates
p=10    # -- network dimension
cluster=2   # -- number of disjoint network components
betaSize=1

#Synthetic_purity(p=p,Rep=Rep,n=n,cluster=cluster,sameModuleN=1,
#dataseed=1,topology="Power", betaSize=betaSize)
```

Index