Package ‘NTW’

March 23, 2017

Type Package

Title Predict gene network using an Ordinary Differential Equation (ODE) based method

Depends R (>= 2.3.0)

Imports mvtnorm, stats, utils

Version 1.24.0

Date 2011-10-8

Author Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini

Maintainer Yuanhua Liu <liuyuanhua@picb.ac.cn>

Description This package predicts the gene-gene interaction network and identifies the direct transcriptional targets of the perturbation using an ODE (Ordinary Differential Equation) based method.

License GPL-2

biocViews Preprocessing

LazyLoad yes

NeedsCompilation no

R topics documented:

NTW-package ................................................................. 2
A.estimation.Srow ...................................................... 3
AP.estimation.Srow ...................................................... 4
comb.matrix ............................................................. 5
methods.regression ..................................................... 6
NTW ................................................................. 7
P.preestimation .......................................................... 8
patterns.priorA .......................................................... 9
sos.data .............................................................. 10

Index 11
**NTW-package**

**Description**

This package includes the functions for estimating the gene-gene interaction network (a matrix, named $A$, with genes as rows and columns) and the associated transcriptional targets of the perturbations (a matrix, named $P$, with genes as rows and perturbations as columns). These estimations are computed with the NTW algorithm, a gene network inference algorithm based on ODE (ordinary differential equation) method, see reference. In this package, the whole $A$ matrix and $P$ matrix are estimated row by row with the function `AP.estimation.Srow`, and built together with the function `NTW`. `AP.estimation.Srow` can be used independently so that estimation of each row can be performed in parallel, improving computation time. For solving the steady state ODE equations, 3 regression methods are supplied: `geo`, `sse` and `ml`, see details in the the corresponding function help pages. In addition, in order to accelerate the estimation of matrix $A$, an option is available to make use of some prior information such as gene association (output from other gene network inference algorithms, or from literature) in `NTW`. The regression methods used in forward or backward mode makes 6 possibilities available for estimating a single row of $A$ matrix. The main functions in this package are listed below,

- **NTW**, to estimate the whole matrix $A$ and $P$ (if $P$ is unknown).
- **AP.estimation.Srow**, to estimate one single row in $A$ and $P$.
- **A.estimation.Srow**, to estimate one single row in $A$ with $P$ known.
- **backward** and **forward**, to estimate one single row of matrix $A$ with different patterns of using prior gene association information.
- **method.geo**, **method.sse** and **method.ml**, to estimate one single row of matrix $A$ with different regression methods.
- **comb.matrix**, sub-function to create all the combinations for regressor locations.
- **P.preestimation**, pre-estimate $P$ matrix according to the gene expression data.

**Details**

<table>
<thead>
<tr>
<th>Package</th>
<th>NTW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>0.99.0</td>
</tr>
<tr>
<td>Date</td>
<td>2010-5-11</td>
</tr>
<tr>
<td>License</td>
<td>GPL-2</td>
</tr>
<tr>
<td>LazyLoad</td>
<td>yes</td>
</tr>
</tbody>
</table>

**Author(s)**

Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini

Maintainer: Yuanhua Liu <liuyuanhua@picb.ac.cn>
A.estimation.Srow

References

Applied method for the inference of gene networks: the bifidobacterium case. to be submitted

---

**A.estimation.Srow**  
*Estimation of a single row in matrix A with the perturbation targets matrix P known*

**Description**

Estimating a single row of gene interaction matrix A when the perturbation targets matrix P is given. The single row in A is then regressed according to the equation AX=P with one of the three regression methods, *geo*, *sse* and *ml*.

**Usage**

```r
A.estimation.Srow(r, cMM.corrected, pred.net, X, P.known, topD, restK, cFlag, sup.drop, noiseLevel)
```

**Arguments**

- `r`: A number indicating the row of A to be estimated.
- `cMM.corrected`: A flag to indicate whether a prior network is applied.
- `pred.net`: A matrix with the same dimensions of A for prior network, which should be specified if cMM.corrected is 1, default is NULL.
- `X`: Gene expression data, a matrix with genes as rows and perturbations as columns.
- `P.known`: A known P matrix with the same dimensions of X.
- `topD`: A parameter in NTW algorithm for keeping the top `topD` combinations of non-zero regressors of row `r` in A, see vignette for details.
- `restK`: A vector (length equals to `nrow(A)`) with each element to indicate the number of non-zero regressors in the corresponding row of A.
- `cFlag`: A flag to tell the regression methods, "geo" for geometric mean method, "sse" for sum of square method and "ml" for maximum likelihood method.
- `sup.drop`: An indication to identify the pattern for using the prior gene association information. 1 for "forward" pattern and -1 for "backward" pattern, see vignette for details.
- `noiseLevel`: Only used in "ml" method, to indicate the noise level in each perturbed experiment.

**Value**

- `A.row`: A vector of estimated row `r` in A.

**Author(s)**

Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini
Examples

## single row estimation without prior gene association information, regression is done by "sse"

data(sos.data)
X<-sos.data
X<-as.matrix(X)
P.known<-matrix(round(runif(nrow(X)*ncol(X), min=0, max=1)), nrow(X), ncol(X))
restK=rep(ncol(X)-1, nrow(X))
topD = round(0.6*nrow(X))
topK = round(0.5*nrow(X))
result<-A.estimation.Srow(r=1,cMM.corrected = 0, pred.net= NULL, X, P.known, topD, restK,
cFlag="sse",sup.drop = -1, noiseLevel=0.1)
result$A.row

## single row estimation with prior gene association information, regression is done by "geo"

pred.net<-matrix(round(runif(nrow(X)*nrow(X), min=0, max=1)), nrow(X), ncol(X))
result<-A.estimation.Srow(r=1,cMM.corrected = 1, pred.net, X, P.known, topD, restK,
cFlag="geo",sup.drop = -1, noiseLevel=0.1)
result$A.row

AP.estimation.Srow

Estimation of a single row in gene interaction matrix A and perturbation targets matrix P

Description

Estimating a single row of gene interaction matrix A and identifying the perturbations which target the corresponding gene of the row. For perturbation identifications, multiple perturbations are considered for one target (gene). Combinations of perturbations are first assumed fixed. The single row in A are then regressed according to equation AX=P with one of the three regression methods, geo, sse and ml. All these combinations will finally be optimized according to the difference between the predicted X with the estimated A and P and the experimental values.

Usage

AP.estimation.Srow(r, cMM.corrected, pred.net, X, IX, topD, restK, cFlag, sup.drop, numP, noiseLevel)

Arguments

r  A number indicating the row of A and P to be estimated.
cMM.corrected  A flag to indicate whether a prior network is applied.
pred.net  A matrix with the same dimensions of A for the prior network, which should be specified if cMM.corrected is 1, default is NULL.
X  Gene expression data, a matrix with genes as rows and perturbations as columns.
IX  The pre-estimated P matrix according to the gene expression data X, with the same dimensions of X or P.
topD  A parameter in NTW algorithm for keeping the top topD combinations of non-zero regressors of row r in A, see vignette for details.
restK  A vector (length equals to nrow(A)) with each element to indicate the number of non-zero regressors in the corresponding row of A.
**comb.matrix**

- **cFlag**: A flag to identify the regression methods, "geo" for geometric mean method, "sse" for sum of square method and "ml" for maximum likelihood method.
- **sup.drop**: A flag to show the pattern for using the prior gene association information. 1 for "forward" pattern and -1 for "backward" pattern.
- **numP**: A number set to limit the possibilities that one gene will be targeted by perturbations. That is at most numP perturbations can directly perturb one gene.
- **noiseLevel**: Only used in "ml" method, to indicate the noise level in each perturbed experiment.

**Value**

- **A.row**: Estimation of the row r in A.
- **P.index**: A vector to show which perturbations target the gene corresponding to row r.

**Author(s)**

Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini

**Examples**

```r
##single row estimation without prior gene association information, regression is done by "sse"
data(sos.data)
X<-sos.data
X<-as.matrix(X)
IX<-P.preestimation(X, topK= round(2*nrow(X)))
restK=rep(ncol(X)-1, nrow(X))
topD = round(0.6*nrow(X))
topK = round(0.5*nrow(X))
umP = round(0.25*nrow(X))
result<-AP.estimation.Srow(r=1,cMM.corrected = 0, pred.net= NULL,X, IX,topD, restK,
cFlag="sse",sup.drop = -1, numP, noiseLevel=0.1)
result$A.row
result$P.index

##single row estimation with prior gene association information, regression is done by "geo"
pred.net<-matrix(round(runif(nrow(X)*nrow(X), min=0, max=1)), nrow(X), ncol(X))
result<-AP.estimation.Srow(r=1,cMM.corrected = 1, pred.net,X, IX,topD, restK,
cFlag="geo",sup.drop = -1, numP, noiseLevel=0.1)
result$A.row
result$P.index
```

---

**comb.matrix**

Create all combinations of vectors

**Description**

Create all combinations of vectors especially for matrices.

**Usage**

```r
comb.matrix(x, y)
```
Methods

Arguments

x  A vector.
y  A vector.

Value

A matrix with nrow(x)*nrow(y) rows and ncol(x)+ncol(y) columns.

Author(s)

Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini

Examples

###A matrix with only one row is obtained###
x<-c(1,2,3)
y<-c(4,5)
comb.matrix(x,y)

###A matrix with 2 rows and 4 columns is obtained###
x<-matrix(x,1,)
y<-matrix(y,,1)
comb.matrix(x,y)

Description

Regression methods to estimate a single row in a gene interaction network (A) with perturbations (P) fixed. These methods differ at the regression criterions: the objective function of geo is the geometric mean, sum of square for sse and maximum likelihood for ml, see vignette for the details.

Usage

method.geo(index.vars, X, pert)
method.sse(index.vars, X, pert)
method.ml(index.vars, X, pert, noiseLevel)

Arguments

index.vars  A vector to select the rows in X on which regression will be used. It sets a group of combination of rows in X.
X  Gene expression data, a matrix with genes as rows and perturbations as columns.
pert  A vector of row r in P.
noiseLevel  Indicate the noise level in each perturbed experiment.

Value

sol  A vector of regression result
error  The result of the objective function.
Estimation of gene interaction matrix $A$ and perturbation targets matrix $P$

Description

This function is used to estimate the whole gene interaction matrix $A$ and the perturbation targets matrix $P$, row-wise, using the NTW algorithm (see reference), based on ODE method. In this method, the linearized ODE can solved using 3 regression methods: `geo`, `sse` and `ml`. In order to save computation time, and improve results, NTW offers the opportunity to input gene association information output from other algorithms or from the literature. The non-null regressors in the gene association network will help fix the regressors to be estimated in the final matrix $A$. Two ways are supplied to use the non-zero information, namely forward and backward approaches. In the “backward” pattern, only the non-zero positions in the prior gene association network will be used as regressors in $A$. While in the “forward” pattern, both these non-zero positions and some other possible positions (depending on $restK$) in $A$ are used as regressors.

Usage

```r
NTW(X, restK, topD, topK = NULL, P.known = NULL, cFlag, pred.net = NULL, sup.drop = -1, numP = NULL, noiseLevel = 0.1)
```

Arguments

- **X**: Gene expression data, a matrix with genes as rows and perturbations as columns.
- **restK**: A vector (length equals to nrow(A)) with each element to indicate the number of non-zero regressors in the corresponding row of $A$.
- **topD**: A parameter in NTW algorithm for keeping the top topD combinations of non-zero regressors of a single row in $A$, see vignette for details.
- **topK**: The number of possible targets of the perturbations, used for pre-estimate the perturbation targets matrix $P$.
- **P.known**: A known $P$ matrix with the same dimensions of $X$.
- **cFlag**: A flag to tell the regression methods, "geo" for geometric mean method, "sse" for sum of square method and "ml" for maximum likelihood method.
- **pred.net**: A matrix with the same dimensions of $A$ for the prior gene association information. Default is NULL.
- **sup.drop**: An indication to show the pattern for using the prior gene association information. $I$ for "forward" pattern and -$I$ for "backward" pattern.
- **numP**: A number set to limit the possibilities that one gene will be targeted by perturbations. That is at most numP perturbations can directly perturb one gene.
- **noiseLevel**: Only used in `ml` method, to indicate the noise level in each perturbed experiment.

Value

- **est.A**: Estimated gene interaction matrix $A$, with genes as rows and columns.
- **est.P**: Estimated perturbation targets matrix $P$, with genes as rows and perturbations as columns.
Pre-estimation of the transcriptional perturbation targets matrix $P$

Description

Pre-estimate the potential transcriptional perturbation targets matrix $P$ according to gene expression data $X$. Those genes with the changes in top $\text{topK}$ will be assumed as possible targets of the perturbations.

Usage

```r
P.preestimation(X, topK)
```

Arguments

- **X**: Gene expression data, a matrix with rows as genes and columns as experiments.
- **topK**: The number of possible targets of the perturbations.

Value

A matrix with the same structure of $X$ or $P$. 

## Examples

### NTW testing without prior gene association information, regression is done by "sse"
```r
data(sos.data)
X <- sos.data
X <- as.matrix(X)
restK <- rep(ncol(X) - 1, nrow(X))
topD = round(0.6 * nrow(X))
topK = round(0.5 * nrow(X))
numP = round(0.25 * nrow(X))
result <- NTW(X, restK, topD, topK, P.known = NULL, cFlag = "sse",
              pred.net = NULL, sup.drop = -1, numP, noiseLevel = 0.1)
result$est.A
result$est.P
```

### NTW testing with prior gene association information, regression is done by "geo"
```r
pred.net <- matrix(round(runif(nrow(X) * nrow(X), min = 0, max = 1)), nrow(X), nrow(X))
result <- NTW(X, restK, topD, topK, P.known = NULL, cFlag = "geo",
              pred.net, sup.drop = -1, numP, noiseLevel = 0.1)
result$est.A
result$est.P
```
**patterns.priorA**

**Author(s)**

Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini

**Examples**

```r
data(sos.data)
X<-sos.data
X<-as.matrix(X)
IX<-P.preestimation(X, topK= round(0.6*nrow(X)))
IX
```

**Description**

Given some information of the gene interaction network, for example the estimated gene association matrix `pred.net` by other algorithms or from literature, NTW can use this prior information to enhance accuracy. NTW offers two approaches to infer the gene network, i.e. forward and backward, on the base of `pred.net`. The former computes further edges than the ones in `pred.net`, while the latter prunes edges.

**Usage**

```r
backward(r, X, pert, topD, restk, cFlag, TA, noiseLevel)
forward(r, X, pert, topD, restk, cFlag, TA, noiseLevel)
```

**Arguments**

- `r`: A number to indicate the row of `A` to be estimated when row `r` of `P` is fixed.
- `X`: Gene expression data, a matrix with genes as rows and perturbations as columns.
- `pert`: Row `r` in `P`.
- `topD`: A parameter in NTW algorithm for keeping the top `topD` combinations of non-zero regressors of row `r` in `A`, see vignette for details.
- `restk`: The number of non-zero regressors for the estimation of row `r` in `A`.
- `cFlag`: A flag to identify the regression methods, ”geo” for geometric mean method, ”sse” for sum of square method and ”ml” for maximum likelihood method.
- `TA`: A vector including the indexes of non-zero elements in row `r` of the network containing a priori information, `pred.net`.
- `noiseLevel`: Only used in ”ml” method, to indicate the noise level in each perturbed experiment.

**Value**

- `A.row`: A vector of estimated row `r` in `A`.
- `CrtValue`: The minimum value from the objective function.

**Author(s)**

Wei Xiao, Yin Jin, Darong Lai, Xinyi Yang, Yuanhua Liu, Christine Nardini
sos.data

**SOS pathway perturbation data**

**Description**
RT-PCR data with 9 genes in SOS pathway of *Escherichia coli* perturbed. These 9 perturbed genes are observed in the RT-PCR experiments.

**Usage**
data(sos.data)

**Format**
`sos.data` is a data frame containing 9 rows for observed genes and 9 columns for perturbations.

**References**

**Examples**
data(sos.data)
sos.data
Index

*Topic arith
  A.estimation.Srow, 3
  AP.estimation.Srow, 4
  comb.matrix, 5
  NTW, 7
  P.preestimation, 8

*Topic datasets
  sos.data, 10

*Topic multivariate
  methods.regression, 6
  patterns.priorA, 9

*Topic package
  NTW-package, 2

A.estimation.Srow, 3
AP.estimation.Srow, 4

backward (patterns.priorA), 9

comb.matrix, 5

forward (patterns.priorA), 9

method.geo (methods.regression), 6
method.ml (methods.regression), 6
method.sse (methods.regression), 6
methods.regression, 6

NTW, 7
NTW-package, 2

P.preestimation, 8
patterns.priorA, 9

sos.data, 10