Package 'iRafNet'

October 26, 2016

Version 1.1-1

Date 2016-10-26

Title Integrative Random Forest for Gene Regulatory Network Inference

Depends R (>= 3.0.0)

Imports ROCR

Description Provides a flexible integrative algorithm that allows information from prior data, such as protein protein interactions and gene knock-down, to be jointly considered for gene regulatory network inference.

License GPL (>= 2)

URL https://www.r-project.org

NeedsCompilation yes

Author Francesca Petralia [aut, cre],
Pei Wang [aut],
Zhidong Tu [aut],
Jialiang Yang [aut],
Adele Cutler [ctb],
Leo Breiman [ctb],
Andy Liaw [ctb],
Matthew Wiener [ctb]

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

Repository CRAN

Date/Publication 2016-10-26 10:37:49

R topics documented:

iRafNet *Integrative random forest for gene regulatory network inference*

Description

This function fits iRafNet, a flexible unified integrative algorithm that allows information from prior data, such as protein-protein interactions and gene knock-down, to be jointly considered for gene regulatory network inference. This function takes as input only one set of sampling scores, computed considering one prior data such as protein-protein interactions or gene expression from knock-out experiments. Note that some of the functions utilized are a modified version of functions contained in the R package randomForest (A. Liaw and M. Wiener, 2002).

Usage

```
iRafNet(X, W, ntree, mtry,genes.name)
```

Arguments

`X`	`(n x p)` Matrix containing expression levels for `n` samples and `p` genes.
`W`	`(p x p)` Matrix containing iRafNet sampling scores. Element `(i,j)` contains score for regulatory relationship `(i -> j)`. Scores must be non-negative. Larger value of sampling score corresponds to higher likelihood of gene `i` regulating gene `j`. Columns and rows of `W` must be in the same order as the columns of `X`. Sampling scores `W` are computed considering one prior data such as protein-protein interactions or gene expression from knock-out experiments.
`ntree`	Numeric value: number of trees.
`mtry`	Numeric value: number of potential regulators to be sampled at each tree node.
`genes.name`	Vector containing gene names. The order needs to match the columns of `X`.

Value

Importance score for each regulatory relationship. The first column contains gene name of regulators, the second column contains gene name of targets, and third column contains corresponding importance scores.

References

Petralia, F., Wang, P., Yang, J., Tu, Z. (2015) Integrative random forest for gene regulatory network inference, *Bioinformatics*, **31**, i197-i205.

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Examples

```
# --- Generate data sets
n<-20                  # sample size
p<-5                   # number of genes
genes.name<-paste("G",seq(1,p),sep="")   # genes name
data<-matrix(rnorm(p*n),n,p)      # generate expression matrix
W<-abs(matrix(rnorm(p*p),p,p))    # generate weights for regulatory relationships

# --- Standardize variables to mean 0 and variance 1
data <- (apply(data, 2, function(x) { (x - mean(x)) / sd(x) } ))

# --- Run iRafNet and obtain importance score of regulatory relationships
out<-iRafNet(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name)
```

`iRafNet_network` *Compute permutation-based FDR of importance scores and return estimated regulations.*

Description

This function computes permutation-based FDR of importance scores and returns gene-gene regulations.

Usage

```
iRafNet_network(out.iRafNet,out.perm,TH)
```

Arguments

`out.iRafNet`	Output object from function `iRafNet`.
`out.perm`	Output object from function `Run_permutation`.
`TH`	Threshold for FDR.

Value

List of estimated regulations.

References

Petralia, F., Song, W.M., Tu, Z. and Wang, P. (2016). New method for joint network analysis reveals common and different coexpression patterns among genes and proteins in breast cancer. *Journal of proteome research*, **15**(3), pp.743-754.

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Xie, Y., Pan, W. and Khodursky, A.B., 2005. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics*, **21**(23), pp.4280-4288.

Examples

```
# --- Generate data sets
n<-20           # sample size
p<-5            # number of genes
genes.name<-paste("G",seq(1,p),sep="")   # genes name
M=5;            # number of permutations
data<-matrix(rnorm(p*n),n,p)       # generate gene expression matrix
data[,1]<-data[,2]                 # var 1 and var 2 interact
W<-abs(matrix(rnorm(p*p),p,p))     # generate weights for regulatory relationships

# --- Standardize variables to mean 0 and variance 1
data <- (apply(data, 2, function(x) { (x - mean(x)) / sd(x) } ))

# --- Run iRafNet and obtain importance score of regulatory relationships
out.iRafNet<-iRafNet(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name)

# --- Run iRafNet for M permuted data sets
out.perm<-Run_permutation(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name,M)

# --- Derive final networks
final.net<-iRafNet_network(out.iRafNet,out.perm,0.001)
```

`iRafNet_permutation` *Derive importance scores for one permuted data.*

Description

This function computes importance score for one permuted data set. Sample labels of target genes are randomly permuted and iRafNet is implemented. Resulting importance scores can be used to derive an estimate of FDR.

Usage

```
iRafNet_permutation(X, W, ntree, mtry,genes.name,perm)
```

Arguments

`X`	`(n x p)` Matrix containing expression levels for `n` samples and `p` genes.
`W`	`(p x p)` Matrix containing iRafNet sampling scores. Element `(i,j)` contains score for regulatory relationship `(i -> j)`. Scores must be non-negative. Larger value of sampling score corresponds to higher likelihood of gene `i` regulating gene `j`. Columns and rows of `W` must be in the same order as the columns of `X`. Sampling scores `W` are computed considering one prior data such as protein-protein interactions or gene expression from knock-out experiments.
`ntree`	Numeric value: number of trees.
`mtry`	Numeric value: number of predictors to be sampled at each node.

genes.name	Vector containing genes name. The order needs to match the rows of x_j.
perm	Integer: seed for permutation.

Value

A vector containing importance score for permuted data.

References

Petralia, F., Wang, P., Yang, J., Tu, Z. (2015) Integrative random forest for gene regulatory network inference, *Bioinformatics*, **31**, i197-i205.

Petralia, F., Song, W.M., Tu, Z. and Wang, P. (2016). New method for joint network analysis reveals common and different coexpression patterns among genes and proteins in breast cancer. *Journal of proteome research*, **15**(3), pp.743-754.

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Examples

```
# --- Generate data sets
n<-20                  # sample size
p<-5                   # number of genes
genes.name<-paste("G",seq(1,p),sep="")   # genes name
data<-matrix(rnorm(p*n),n,p)       # generate expression matrix
W<-abs(matrix(rnorm(p*p),p,p))     # generate weights for regulatory relationships

# --- Standardize variables to mean 0 and variance 1
data <- (apply(data, 2, function(x) { (x - mean(x)) / sd(x) } ))

# --- Run iRafNet and obtain importance score of regulatory relationships
out.iRafNet<-iRafNet(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name)

# --- Run iRafNet for one permuted data set and obtain importance scores
out.perm<-iRafNet_permutation(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name,perm=1)
```

---

roc_curve *Plot receiver operating characteristic (ROC) curve for weighted network generated by iRafNet*

Description

This function uses R package ROCR to plot ROC curves for iRafNet object.

Usage

```
roc_curve(out, truth)
```

Arguments

out	Output from iRafNet.
truth	Matrix of true regulations. Rows correspond to different regulations and match rows of out. First column contains name of regulators, second column contains name of targets and third column contains a binary variable equal 1 in case of regulation and 0 otherwise.

Value

Plot ROC curve and return area under ROC curve.

References

Petralia, F., Wang, P., Yang, J., Tu, Z. (2015) Integrative random forest for gene regulatory network inference, *Bioinformatics*, **31**, i197-i205.

Sing, Tobias, et al. (2005) ROCR: visualizing classifier performance in R, *Bioinformatics*, **21**, 3940-3941.

Examples

```
# --- Generate data sets
n<-20                  # sample size
p<-5                   # number of genes
genes.name<-paste("G",seq(1,p),sep="")   # genes name
data<-matrix(rnorm(p*n),n,p)    # generate expression matrix
data[,1]<-data[,2]              # var 1 and 2 interact
W<-abs(matrix(rnorm(p*p),p,p))  # generate score for regulatory relationships

# --- Standardize variables to mean 0 and variance 1
data <- (apply(data, 2, function(x) { (x - mean(x)) / sd(x) } ))

# --- Run iRafNet and obtain importance score of regulatory relationships
out<-iRafNet(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name)

# --- Matrix of true regulations
truth<-out[,seq(1,2)]
truth<-cbind(as.character(truth[,1]),as.character(truth[,2])
,as.data.frame(rep(0,,dim(out)[1])));
truth[(truth[,1]=="G2" & truth[,2]=="G1") | (truth[,1]=="G1" & truth[,2]=="G2"),3]<-1

# --- Plot ROC curve and compute AUC
auc<-roc_curve(out,truth)
```

Run_permutation *Derive importance scores for M permuted data sets.*

Description

This function computes importance score for `M` permuted data sets. Sample labels of target genes are randomly permuted and iRafNet is implemented. Resulting importance scores can be used to derive an estimate of FDR.

Usage

```
Run_permutation(X, W, ntree, mtry,genes.name,M)
```

Arguments

`X`	`(n x p)` Matrix containing expression levels for `n` samples and `p` genes.
`W`	`(p x p)` Matrix containing iRafNet sampling scores. Element `(i,j)` contains score for regulatory relationship `(i -> j)`. Scores must be non-negative. Larger value of sampling score corresponds to higher likelihood of gene `i` regulating gene `j`. Columns and rows of `W` must be in the same order as the columns of `X`. Sampling scores `W` are computed considering one prior data such as protein-protein interactions or gene expression from knock-out experiments.
`ntree`	Numeric value: number of trees.
`mtry`	Numeric value: number of predictors to be sampled at each node.
`genes.name`	Vector containing genes name. The order needs to match the rows of `x_j`.
`M`	Integer: total number of permutations.

Value

A matrix with `I` rows and `M` columns with `I` being the total number of regulations and `M` the number of permutations. Element `(i,j)` corresponds to the importance score of interaction `i` for permuted data `j`.

References

Petralia, F., Wang, P., Yang, J., Tu, Z. (2015) Integrative random forest for gene regulatory network inference, *Bioinformatics*, **31**, i197-i205.

Petralia, F., Song, W.M., Tu, Z. and Wang, P. (2016). New method for joint network analysis reveals common and different coexpression patterns among genes and proteins in breast cancer. *Journal of proteome research*, **15**(3), pp.743-754.

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. *R News* **2**, 18–22.

Examples

```
# --- Generate data sets
n<-20                  # sample size
p<-5                   # number of genes
genes.name<-paste("G",seq(1,p),sep="")   # genes name
M=5;            # number of permutations

data<-matrix(rnorm(p*n),n,p)       # generate expression matrix
W<-abs(matrix(rnorm(p*p),p,p))          # generate score for regulatory relationships

# --- Standardize variables to mean 0 and variance 1
data <- (apply(data, 2, function(x) { (x - mean(x)) / sd(x) } ))

# --- Run iRafNet and obtain importance score of regulatory relationships
out.iRafNet<-iRafNet(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name)

# --- Run iRafNet for M permuted data sets
out.perm<-Run_permutation(data,W,mtry=round(sqrt(p-1)),ntree=1000,genes.name,M)
```

Index