Package 'bc3net'

February 19, 2015

Version 1.0.2

Title BC3NET

Date 2013-04-04

Author Ricardo de Matos Simoes and Frank Emmert-Streib

Maintainer Ricardo de Matos Simoes <r.dematossimoes@qub.ac.uk>

Depends R (>= 2.10.0), c3net, infotheo, igraph, Matrix, lattice

Suggests

Description This package implements the BC3NET algorithm for gene regulatory network inference (de Matos Simoes and Frank Emmert-Streib, Bagging Statistical Network Inference from Large-Scale Gene Expression Data, PLoS ONE 7(3): e33624)

License GPL (>= 2)

URL http://www.bio-complexity.com

NeedsCompilation no

Repository CRAN

Date/Publication 2013-04-07 15:01:54

R topics documented:

bc3net-package *BC3NET Gene Regulatory network Inference*

Description

The basic idea of BC3NET is to generate from one dataset D_s, consisting of s samples, an ensemble of B independent bootstrap datasets D_k by sampling from D(s) with replacement by using a non-parametric bootstrap (Efron 1993). Then, for each generated data set D_k in the ensemble, a network G^b_k is inferred by using C3NET (Altay 2010a). From the ensemble of networks G^b_k we construct one weighted network G^b_w which is used to determine the statistical significance of the connection between gene pairs. This results in the final binary, undirected network G.

A base component of BC3NET is the inference method C3NET introduced in Altay (2010a), which we present in the following in a modified form to obtain a more efficient implementation. Briefly, C3NET consists of three main steps. First, mutual information values among all gene pairs are estimated. Second, an extremal selection strategy is applied allowing each of the p genes in a given dataset to contribute at most one edge to the inferred network. That means we need to test only p different hypotheses and not p(p-1)/2. This potential edge corresponds to the hypothesis test that needs to be conducted for each of the p genes. Third, a multiple testing procedure is applied to control the type one error. In the above described context, this results in a network G^b_k.

Details

Package:	bc3net
Type:	Package
Version:	1.0.0
Date:	2012-01-12
License:	GPL (>=2)

bc3net.R c3mtc.R makenull.R mimwrap.R getpval.R mat2igraph.R

Author(s)

Ricardo de Matos Simoes <r.dematossimoes@qub.ac.uk> Frank Emmert-Streib <f.emmert-streib@qub.ac.uk>
Maintainer: Ricardo de Matos Simoes <r.dematossimoes@qub.ac.uk>

References

de Matos Simoes R, Emmert-Streib F., Bagging statistical network inference from large-scale gene expression data., PLoS One. 2012;7(3):e33624. Epub 2012 Mar 30.

See Also

C3NET, MINET, INFOTHEO

Examples

```
data(expmat)
bnet=bc3net(expmat)

data(expmat)
cnet=c3mtc(expmat)
```

---

bc3net *Bc3net gene regulatory network inference*

Description

The basic idea of BC3NET is to generate from one dataset D_s, consisting of s samples, an ensemble of B independent bootstrap datasets D_k by sampling from D(s) with replacement by using a non-parametric bootstrap (Efron 1993). Then, for each generated data set D_k in the ensemble, a network G^b_k is inferred by using C3NET (Altay 2010a). From the ensemble of networks G^b_k we construct one weighted network G^b_w which is used to determine the statistical significance of the connection between gene pairs. This results in the final binary, undirected network G.

A base component of BC3NET is the inference method C3NET introduced in Altay (2010a), which we present in the following in a modified form to obtain a more efficient implementation. Briefly, C3NET consists of three main steps. First, mutual information values among all gene pairs are estimated. Second, an extremal selection strategy is applied allowing each of the p genes in a given dataset to contribute at most one edge to the inferred network. That means we need to test only p different hypotheses and not p(p-1)/2. This potential edge corresponds to the hypothesis test that needs to be conducted for each of the p genes. Third, a multiple testing procedure is applied to control the type one error. In the above described context, this results in a network G^b_k.

Usage

```
bc3net(dataset, boot=100, estimator="pearson", disc="equalwidth", mtc1=TRUE,
alpha1=0.05, nullit=NA, null=c(), adj1="bonferroni", mtc2=TRUE,
alpha2=0.05, adj2="bonferroni",
weighted=TRUE, igraph=TRUE, verbose=FALSE)
```

Arguments

dataset	gene expression dataset where rows define genes and columns samples
boot	default 100 bootstrap datasets are generated to infer an ensemble of c3net gene regulatory networks
estimator	estimators for continuous variables "pearson", "spearman", "kendall", "spearman"
	estimators for discrete variables "emp", "mm","sg","shrink"
disc	required for discrete estimators, method for discretize function (see infotheo package) "equalwidth" (default), "equalfreq", "globalequalwidth"

nullit	nullit defines the size of the generated null distribution vector used for hypothesis testing of significant edges inferred by c3net. The null distribution of mutual information is generated from sample and gene label randomization. number of iterations, where the default is defined by nullit=ceiling(10^5/(((genes*genes)/2)-genes)) genes: number of genes
null	assign alternatively an external null distribution vector
mtc1	consider multiple hypothesis testing for edges inferred by c3net
alpha1	significance level for mtc1
adj1	if mtc1==TRUE default multiple hypothesis testing procedure for c3net inferred edges using "bonferroni" (default) alternatively use "holm", "hochberg", "hommel", "bonferroni", "BH", "BY","fdr", "none" (see ?p.adjust())
mtc2	Consider multiple hypothesis testing for edges inferred by bc3net. A binomial test is performed for each gene pair with an ensemble consensus rate >0 consider multiple hypothesis testing for edges inferred by bc3net
alpha2	significance level for mtc2
adj2	Consider multiple hypothesis testing for edges inferred by bc3net. if mtc2==TRUE "bonferroni" is used as multiple hypothesis testing procedure. alternatively use "holm", "hochberg", "hommel", "bonferroni", "BH", "BY","fdr", "none"
weighted	A weighted network is returned, where the weights denote the ensemble consensus rate of bc3net.
igraph	A bc3net igraph object is returned.
verbose	Return processing information of running procedures.

Details

BC3NET Gene regulatory network inference

Value

'bc3net' returns a gene regulatory network formated as adjacency matrix, as weighted matrix where the edge weights are defined by the corresponding mutual information values or as undirected weighted or unweighted igraph object.

Author(s)

de Matos Simoes R, Emmert-Streib F.

References

Altay G, Emmert-Streib F. Inferring the conservative causal core of gene regulatory networks. BMC Syst Biol. 2010 Sep 28;4:132.

de Matos Simoes R, Emmert-Streib F. Bagging statistical network inference from large-scale gene expression data. PLoS One. 2012;7(3):e33624. Epub 2012 Mar 30.

de Matos Simoes R, Emmert-Streib F. Influence of statistical estimators of mutual information and data heterogeneity on the inference of gene regulatory networks. PLoS One. 2011;6(12):e29279. Epub 2011 Dec 29.

See Also

C3NET c3mtc

Examples

```
data(expmat)
bnet=bc3net(expmat)
```

`c3mtc` *'c3mtc' gene regulatory network inference using c3net with multiple testing correction procedure*

Description

We present in the following the inference method C3NET introduced in Altay (2010a) in a modified form to obtain a more efficient implementation. Briefly, C3NET consists of three main steps. First, mutual information values among all gene pairs are estimated. Second, an extremal selection strategy is applied allowing each of the p genes in a given dataset to contribute at most one edge to the inferred network. That means we need to test only p different hypotheses and not p(p-1)/2. This potential edge corresponds to the hypothesis test that needs to be conducted for each of the p genes. Third, a multiple testing procedure is applied to control the type one error.

In order to determine the statistical significance of the mutual information values between genes we test for each pair of genes the following null hypothesis.

H_0^I: The mutual information between gene i and j is zero.

Because we are using a nonparametric test we need to obtain the corresponding null distribution for H_0^I from a randomization of the data.

The formulated null hypothesis is performed by permuting the sample and gene labels for all genes of the entire expression matrix at once. The vector of the mutual information null distribution is obtained from repeated randomizations for a given number of iterations.

Usage

```
c3mtc(dataset, null=NULL, mtc=TRUE, adj="bonferroni", alpha=0.05, nullit=NA,
estimator="pearson", disc="none", adjacency=FALSE, igraph=TRUE)
```

Arguments

dataset	gene expression dataset where rows define genes and columns samples
nullit	nullit defines the size of the generated null distribution vector used for hypothesis testing of significant edges inferred by c3net. The null distribution of mutual information is generated from sample and gene label randomization. default number of iterations: nullit=ceiling(10^5/(((genes*genes)/2)-genes)) genes: number of genes
estimator	minet package (continuous estimators) "pearson", "spearman", "kendall", "spearman" minet package (discrete estimators) "mi.empirical", "mi.mm","mi.sg","mi.shrink" c3net gaussian estimator (pearson) "gaussian" bspline requires installation of "mis_calc" "bspline"
disc	only required for discrete estimators (minet package) "equalfreq", "equalwidth"
mtc	consider multiple hypothesis testing for edges inferred by c3net
adj	if mtc==TRUE default multiple hypothesis testing procedure for c3net inferred edges using "bonferroni" (default) alternatively use "holm", "hochberg", "hommel", "bonferroni", "BH", "BY","fdr", "none" (see ?p.adjust())
alpha	significance level for mtc after multiple hypothesis testing correction
adjacency	return an adjacency matrix
igraph	return igraph object
null	If NULL a null distribution vector is generated from a sample label and gene label permutation of the gene expression matrix. For the ensemble inference of one dataset an external null distribution vector is suggested for decreasing running time.

Value

'c3mtc' returns a gene regulatory network formated as adjacency matrix, as weighted matrix where the edge weights are defined by the corresponding mutual information values or as undirected weighted or unweighted igraph object.

Author(s)

de Matos Simoes R, Emmert-Streib F.

References

Altay G, Emmert-Streib F. Inferring the conservative causal core of gene regulatory networks. BMC Syst Biol. 2010 Sep 28;4:132. PubMed PMID: 20920161; PubMed Central PMCID: PMC2955605.

de Matos Simoes R, Emmert-Streib F. Bagging statistical network inference from large-scale gene expression data. PLoS One. 2012;7(3):e33624. Epub 2012 Mar 30. PubMed PMID: 22479422; PubMed Central PMCID: PMC3316596.

de Matos Simoes R, Emmert-Streib F. Influence of statistical estimators of mutual information and data heterogeneity on the inference of gene regulatory networks. PLoS One. 2011;6(12):e29279. Epub 2011 Dec 29. PubMed PMID: 22242113; PubMed Central PMCID: PMC3248437.

See Also

c3

Examples

```
data(expmat)
net=c3mtc(expmat)
```

---

expmat *Test gene expression dataset*

Description

The dataset is test dataset of a gene expression matrix with 100 genes and 100 samples

Usage

```
data(expmat)
```

Format

A matrix with 100 observations and 100 variables.

References

de Matos Simoes R, Emmert-Streib F., "Bagging statistical network inference from large-scale gene expression data" PLoS One. 2012;7(3):e33624. Epub 2012 Mar 30. PubMed PMID: 22479422; PubMed Central PMCID: PMC3316596.

Examples

```
data(expmat)
```

---

mimwrap *Wrapper function for mutual information matrix estimators*

Description

Mutual information matrix estimation wrapper function for various mutual information estimators. Depends on infotheo package for mutual information estimators on discrete variables.

Usage

```
mimwrap(dataset, estimator="pearson", disc="equalwidth", bins = sqrt(ncol(dataset)))
```

Arguments

dataset	Data gene expression matrix where rows denote genes (features) and columns samples.
estimator	estimators for continuous variables "pearson" (default), "spearman", "kendall", "spearman" estimators for discrete variables (infotheo package) "emp", "mm","sg","shrink"
disc	only required for discrete estimators (see infotheo package) "equalwidth" (default), "globalequalwidth" , "equalfreq"
bins	number of bins for the descretize function (infotheo), default sqrt(ncol(dataset))

Details

A mutual information matrix is estimated from a gene expression data set

Value

mimwrap returns a symmetric mutual information matrix for various mutual information estimators.

References

Patrick E Meyer, Frederic Lafitte and Gianluca Bontempi, minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information, BMC Bioinformatics 2008, 9:461

Carsten O. Daub, Ralf Steuer, Joachim Selbig, and Sebastian Kloska, Estimating mutual information using B-spline functions - an improved similarity measure for analysing gene expression data, BMC Bioinformatics. 2004; 5: 118

de Matos Simoes R, Emmert-Streib F., Bagging statistical network inference from large-scale gene expression data., PLoS One. 2012;7(3):e33624. Epub 2012 Mar 30.

Examples

```
data(expmat)
mim <- mimwrap(expmat)
```

Index