

Package ‘GCnetinf’

August 6, 2015

Type Package

Title GCnetinf - Granger causality tests for network inference

Version 1.0

Date 2015-05-03

Author Miguel Lopes

Maintainer Miguel Lopes <miguelaglopes@gmail.com>

Description Takes as input a multiple time series, returns dynamic causality scores between variables.

Imports tseries, ppcor, randomForest, lars

License GPL-3

R topics documented:

GCnetinf-package	1
gcausal	2
netinf11	4

Index	6
--------------	----------

GCnetinf-package	<i>GCnetinf</i>
------------------	-----------------

Description

This package computes dynamic causality scores between variables from a multiple time series matrix. Consists of two main functions, described below. The first function is "gcausal" and estimates linear Granger causality (GC) scores. Bivariate and conditional (1 order) GC tests are available, returning a matrix of z-scores, where the element [i,j] is the score from variable i to j. The conditional GC score from a cause to effect is the minimum 1 order conditional GC score obtained, for all individual conditioning variables. An heuristic is applied in order to speed up this search. Options for lag selection and integrated variables are also available. A method to identify variables with common causes (siblings) is also implemented. The second function is "netinf11" and implements dynamic versions (1-lag) of state of the art network inference algorithms: bivariate mutual information, aracne, mrmr, cmim, mimr, random forests, and lasso/least angle regression. These methods are described in Lopes 2015 (PhD thesis).

Details

Package: GCnetinf
Type: Package
Version: 1.0
Date: 2015-05-03
License: GPL3

See the entries on gcausal and netinf11 on how to use this package.

Author(s)

Miguel Lopes

Maintainer: Miguel Lopes <miguelaglopes@gmail.com>

References

Miguel Lopes, Inference of gene networks from time series expression data and application to type 1 Diabetes, 2015 (PhD Thesis)

gcausal

Granger causality (GC) tests for multiple time series.

Description

This function estimates bivariate and conditional (1 order) linear Granger causality between each pair of variables in the dataset. Returns a p times p matrix of GC z-scores, obtained with an F-test on the residuals of restricted and unrestricted models. The conditional GC scores are the minimum of first order conditional GC scores (a score designated by GC3 herein). In order to improve its speed when the number of variables is high, an approximation is adopted based on a search heuristic. Its accuracy and speed can be controlled with a user given parameter (see below). It is possible to filter variable pairs which are identified as having common causes (siblings). This method is referred in Lopes 2015 as co-regulation identification, in the context of gene regulations. In this case, scores between siblings are assigned NA values.

GC tests may consider multiple lags and may be corrected to consider integrated variables (Toda-Yamamoto modification). The default option is to consider a single first lag, but this may be modified (see below).

See Lopes 2015 (PhD thesis).

Usage

```
gcausal(datamatrix,  
type = "conditional",  
lagmethod = "first",  
maxnumlags = 1,  
maxlag = 1,  
crit = "aicc",  
ty.test = FALSE,  
stat.pars,
```

```
sibling.filter=FALSE,
sf.mincor=0.7,
sf.maxlag=2 ,
sf.matrix,
gc3rank.method="dynamic",
gc3rank.approx=3,
gc3rank.matrix)
```

Arguments

<code>datamatrix</code>	A numeric matrix of dimension n times p (samples are rows, variables are columns). No NA or Inf values allowed.
<code>type</code>	Character string, either "bivariate" or "conditional" (default).
<code>lagmethod</code>	Character string, either "first" (default) or "fsel". The lag(s) of the target are always the first (its number may be estimated by AIC). Then, an equal number of predictor lags are considered. However, these may be selected in a forward selection procedure, instead of being the first. For instance, the lags returned by "first" may be 1,2,3, and the lags returned by "fsel" 2,1,4.
<code>maxnumlags</code>	Integer (default 1). This parameter defines the maximum number of lags (of a variable) to be included in the GC model.
<code>maxlag</code>	Integer (default 1). This parameter defines the maximum lag to be included in the GC model (it may be different than <code>maxnumlags</code> when <code>lagmethod="fsel"</code>).
<code>crit</code>	Character string, either "aicc" (default), "aic" or "bic". This parameter defines the criterion to assess linear models.
<code>ty.test</code>	Logical, TRUE or FALSE (default FALSE). If TRUE, the Toda-Yamamoto modified GC test is used (deals with integrated variables).
<code>stat.pars</code>	Named list of parameters to estimate the order of integration and assess stationarity. Elements in the list should be named "maxorder", "method" and "cutoff". "maxorder" is integer and defines the maximum order of integration in the integration tests; "method" is a character string, and defines the method to test the null hypothesis of stationarity (either "KPSS" or "adf"); "cutoff" is numeric, and defines the p-value level below which stationarity is rejected. Default is <code>list(maxorder=2, method="KPSS", cutoff=0.05)</code> .
<code>sibling.filter</code>	Logical, TRUE or FALSE (default). If TRUE scores between siblings (variables with common causes) are filtered (the respective scores are NA). Sibling identification is as described in Lopes 2015.
<code>sf.mincor</code>	Numeric (default 0.7). This parameter defines the minimum linear correlation (absolute value) for sibling identification.
<code>sf.maxlag</code>	Integer (default 2). This parameter defines the maximum considered lag in sibling identification.
<code>sf.matrix</code>	(optional) p times p sibling matrix (if already computed). Non zero elements indicate siblings.
<code>gc3rank.method</code>	Character string, either "static" or "dynamic" (default). This parameter defines the method to compute the ranking in GC3.
<code>gc3rank.approx</code>	Integer (default 3). This parameter controls the speed of the GC3 search (parameter "s" in Lopes 2015). The lower the faster, and less precise. Maximum is p (equivalent to the full search).
<code>gc3rank.matrix</code>	(optional) Ranking matrix for GC3 (if already computed)

Value

A p times p matrix of GC z-scores. Score [i,j] is from element i to element j.

Author(s)

Miguel Lopes

References

Miguel Lopes, Inference of gene networks from time series expression data and application to type 1 Diabetes, 2015 (PhD Thesis)

Examples

```
#datamatrix=mat.or.vec(50,20)+rnorm(50*20,0,1)
#gcscores=gcausal(datamatrix, type="bivariate")
```

netinf1l

1-lag dynamic network inference algorithms for time series.

Description

This function implements 1-lag dynamic network inference algorithms for time series. For each target, predictors are lagged (1 lag), and scored according to the selected model. Returns a p times p matrix of scores. Score [i,j] is from variable i to j.

Implemented methods are bivariate mutual information, clr, aracne, mrmr, cmim, mimr, random forests (RF), and lasso/least angle regression (lars). The last two call the packages randomForest and lars.

In RF, variable importance is measured with the mean decrease in node impurity ("IncNodePurity", in the case of regression, measured with RSS). The arguments of the function randomForest may be passed through the argument rf.pars.

In lars, scores are obtained as the average of predictor coefficients for various lambda values. The arguments of the lars functions may be passed through the arguments lars.pars and predict.lars.pars.

See Lopes 2015 (PhD thesis).

Usage

```
netinf1l(datamatrix, Methods, mi.cutoff = 0.05, rf.pars, lars.pars, predict.lars.pars)
```

Arguments

datamatrix	A numeric matrix of dimension n times p (samples are rows, variables are columns). No NA or Inf values allowed.
Methods	Character string, either "mi" (mutual information) "aracne", "clr", "mrmr", "cmim", "mimr", "rf" (random forests), or "lars" (least angle regression). This parameter selects the used inference method. Multiple methods may be selected.
mi.cutoff	Numeric (default 0.05). The mutual information is estimated as a function of the linear correlation (Gaussian assumption). This value is the significance cut-off for the statistical test on linear correlations.

<code>rf.pars</code>	Named list of parameters for the function <code>randomForest::randomForest</code> . Default is <code>list(mtry=sqrt(number of variables), ntree=1000)</code> .
<code>lars.pars</code>	Named list of parameters for the function <code>lars::lars</code> . Default is <code>list(type="lasso", use.gram=TRUE)</code>
<code>predict.lars.pars</code>	Named list of parameters for the function <code>lars::predict.lars</code> . Default is <code>list(type="coefficients", mode="fraction", s=seq(0.1,1,0.01))</code>

Value

A list in which element is a p times p matrix of network scores (relative to a method). Score $[i,j]$ is from element i to element j .

Author(s)

Miguel Lopes

References

Miguel Lopes, Inference of gene networks from time series expression data and application to type 1 Diabetes, 2015 (PhD Thesis)

Index

gcausal, [2](#)
GCnetinf (GCnetinf-package), [1](#)
GCnetinf-package, [1](#)
netinf11, [4](#)