

Package ‘netSEM’

June 12, 2018

Type Package

Title Network Structural Equation Modeling

Description The network structural equation modeling conducts a network statistical analysis on a data frame of coincident observations of multiple continuous variables [1]. It builds a pathway model by exploring a pool of domain knowledge guided candidate statistical relationships between each of the variable pairs, selecting the 'best fit' on the basis of a specific criteria such as adjusted r-squared value.

This work was funded under U. S. Dept. of Energy, Prime Award No. DE-E-0004946, Award Agreement No. 60220829-51077-T.

[1] Bruckman, Laura S., Nicholas R. Wheeler, Junheng Ma, Ethan Wang, Carl K. Wang, Ivan Chou, Jiayang Sun, and Roger H. French. (2013) <doi:10.1109/ACCESS.2013.2267611>.

Version 0.5.0

Depends R (>= 3.1.0)

Imports DiagrammeR (>= 0.9.2), DiagrammeRsvg (>= 0.1), htmlwidgets (>= 1.2), knitr (>= 1.20), magrittr (>= 1.5), MASS (>= 7.3-47), rsvg (>= 1.1), svglite (>= 1.2.1), png (>= 0.1-7), segmented (>= 0.5-3.0), gtools (>= 3.5.0)

License GPL (>= 2)

LazyData true

Suggests rmarkdown, testthat

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Author Wei-Heng Huang [aut, cre] (<<https://orcid.org/0000-0002-6609-4981>>),
Nicholas Wheeler [aut] (<<https://orcid.org/0000-0003-2248-8919>>),
Addison Klinke [aut] (<<https://orcid.org/0000-0002-6985-7657>>),
Yifan Xu [aut] (<<https://orcid.org/0000-0003-1696-0228>>),
Wenyu Du [aut] (<<https://orcid.org/0000-0002-8672-9104>>),
Abdulkerim Gok [aut] (<<https://orcid.org/0000-0003-3433-7106>>),
Devin Gordon [ctb] (<<https://orcid.org/0000-0002-5919-0422>>),
Yu Wang [ctb] (<<https://orcid.org/0000-0003-1353-2578>>),

Jiqi Liu [ctb] (<<https://orcid.org/0000-0003-2016-4160>>),
 Alan Curran [ctb] (<<https://orcid.org/0000-0002-4505-8359>>),
 Justin Fada [ctb] (<<https://orcid.org/0000-0002-0029-5051>>),
 Xuan Ma [ctb] (<<https://orcid.org/0000-0003-2361-2846>>),
 Jennifer Braid [ctb] (<<https://orcid.org/0000-0002-0677-7756>>),
 Laura Bruckman [aut] (<<https://orcid.org/0000-0003-1271-1072>>),
 Roger French [aut, cph] (<<https://orcid.org/0000-0002-6162-0532>>)

Maintainer Wei-Heng Huang <wxh272@case.edu>

Repository CRAN

Date/Publication 2018-06-12 10:33:50 UTC

R topics documented:

acrylic	2
backsheet	3
crack	4
findrelation	5
genInit	6
IVfeature	6
netSEM	7
netSEMm	8
path	9
pathwayPredict	11
pathwayRMSE	12
pet	13
plot.netSEM	14
subsetData	15
summary.netSEM	16
Index	17

acrylic	<i>A data frame of an acrylic degradation experiment</i>
---------	--

Description

The data set is a study of photodegradation of acrylic polymer. In this work, polymeric samples were exposed to different levels of light exposures and resulting optical changes were determined through optical spectroscopy. *IrradTot* (total applied irradiance) is the main exogenous variable and *YI* (yellowness index) is the endogenous variable (response). The other columns in the data set (*IAD1*, *IAD2*, *IAD2p*, and *IAD3*) are induced absorbance to dose values extracted from optical absorbance spectra as single metrics and used as intermediate unit level endogenous (response) variables in the netSEM analysis.

Usage

```
data(acrylic)
```

Format

A 357 by 6 data frame of continuous variables

Source

Solar Durability and Lifetime Extension (SDLE) Research Center, Case Western Reserve University

backsheet

Backsheet PET/PET/EVA Degradation

Description

A backsheet is a polymeric cover of photovoltaic (PV) module and is designed to protect the inner components of module. Typical backsheets consist of three layers of polymers to prohibit diffusion of water and oxygen, as well as to protect human beings from electrical shock. It is critical that your solar panel has a backsheet that is of high quality and can withstand various environmental elements for 25 long years. Polyethylene terephthalate (PET) is an important material, and it mainly used as core and outer layer of backsheets and provide mechanical stability and electrical isolation. However, PET based Backsheets are highly susceptible to moisture and ultraviolet (UV) irradiance. Degradation of backsheets will causes severe economic loss and safty issue. Indoor accelerated exposures are used to study backsheets degradation within short time and predict the performance of backsheets exposed to the real world.

Usage

data(backsheet)

Format

A data frame with 110 rows and 5 variables:

YI Yellowness index of PET outer layer measured at every exposure step

Hours Exposure time of PET in Damp Heat condition

oxidation Formed conjugation structure during oxidation (FTIR peak at 1552 cm-1)

hydrolysis Formed acid and alcohol structure during hydrolysis (FTIR peak 2900 - 3300 cm-1)

crystallization Crystallinity calculated as the ratio of FTIR peak at 1340 cm-1 to 1370 cm-1

Details

The backsheet degradation can be characterized with yellowness index (YI), which represents the color change of polymer and is associated with chemical change due to irradiance, high temperature and other process. The YI value of a PV backsheet also relates to the module efficiency. In addition, the Fourier-transform infrared spectroscopy (FTIR) is a effective method to study the chemical change of polymers. This non-destructive measurement enable people to obtain qualitative information of polymer functional groups and the relative amount of each group in the sample.

In this example, a dataset containing the YI of PET based backsheets (PET/PET/EVA) exposed to 1,500 hours of Damp Heat with no irradiance and an relative humidity of 85% and a chamber temperature of 85 degree celsius. The PET samples were pull out from exposure chamber every 500 hours and the YI measurement was conducted at every step. The stress variable is exposure time with unit of hour. Mechanistic variables from FTIR are included to track chemical changes in the materials related to polymer chain scission due to hydrolysis, crystallinity change and oxidation.

Author(s)

Yu Wang, Wei-heng Huang, Laura S. Bruckman, Roger H. French

Source

Solar Durability and Lifetime Extension (SDLE) Research Center, Case Western Reserve University

crack

Crack Quantification for Photovoltaic Backsheets

Description

A dataset containing the average normalized crack depth for photovoltaic backsheets with inner layers of either ethylene-vinyl acetate or polyethylene exposed to 4,000 hours of continuous UVA irradiance with an intensity of 1.55 w/m² at 340 nm and a chamber temperature of 70 deg C. Mechanistic variables from Fourier transform infrared spectroscopy are included to track chemical changes in the materials related to cracking. See the journal article titled 'A Non- Destructive Method for Crack Quantification in Photovoltaic Backsheets Under Accelerated and Real-World Exposures' in Polymer Degradation and Stability for more details.

Usage

data(crack)

Format

A data frame with 130 rows and 5 variables:

dAvgNorm Average crack depth normalized by the backsheet's inner layer thickness

uva360dose Integrated, cumulative photodose for all wavelengths less than 360 nm

crys730 Percent crystallinity calculated from the ratio of CH₂ rocking peaks at 731 and 720 cm⁻¹

carb1715 Ketone carbonyl index calculated as the ratio of intensities at 1715 and 2851 cm⁻¹

carbPC1 1st principal component score from the carbonyl region (1500-1800 cm⁻¹)

Author(s)

Addison G. Klinke, Abdulkерim Gok, Laura S. Bruckman, Roger H. French

Source

Solar Durability and Lifetime Extension (SDLE) Research Center, Case Western Reserve University

findrelation	<i>Find relationship between any two variables</i>
--------------	--

Description

Find relationship between any two variables.

Usage

```
findrelation(iVar, iResp, criterion = "adj.R2", modelNames, str = FALSE,
             cRes.all, cRes.best, cRes.print, x, nlsInits, nRes)
```

Arguments

iVar	The first parameter.
iResp	The second parameter.
criterion	What model fit parameter is used for ranking? (Default: adj.R2)
modelNames	A vector of model names chosen from "SL", "Quad", "SQquad", "Exp", "Log", "nls", "CP"
str	TRUE/FALSE Is this a 'strength' type model fitting? (whether to use ^2 and sqrt functions)
cRes.all	A dataframe of all results, passed in by the netSEMm function
cRes.best	A dataframe of best results, passed in by the netSEMm function
cRes.print	A dataframe of results to print, passed in by the netSEMm function
x	A dataframe of data values, passed in by the netSEMm function
nlsInits	A vector of nls initialization coefficients, passed in by the netSEMm function
nRes	number of cells in the print variable, value passed in by the netSEMm function

Value

a list of the following items:

- "cRes.all": A dataframe of all results.
- "cRes.best": A dataframe of best results.
- "cRes.print": A dataframe of results to print

genInit	<i>Generate initial values for nls function</i>
---------	---

Description

Generate multiple initial vectors for the nls function in netSEM().

Usage

```
genInit(bounds = list(a1 = c(-3, 3), a2 = c(-3, 3), a3 = c(-3, 3)), k = 50)
```

Arguments

bounds	a list of three vectors of length = 2. Each vector gives the upper and lower limits of an interval from which the initial values are randomly generated. The default values list(a1 = c(-3, 3), a2 = c(-3, 3), a3 = c(-3, 3)) sets limits of all three initial values to be (-3, 3).
k	a positive integer (default = 50). The number of initial vectors to generate.

Details

Currently the non-linearizable function included in netSEM() is $y = a + b * \exp(c * x)$, where a, b and c are coefficients to be estimated. Thus an initial vector contains three values. The random initial values are generated by a uniform distribution between the bounds.

Value

a data frame. Each column corresponds to a coefficient. Each row corresponds to a random initial vector.

Examples

```
genInit(list(a1 = c(0,2), a2 = c(4,5), a3 = c(-1, -0.5)), k = 20 )
```

IVfeature	<i>IV features data</i>
-----------	-------------------------

Description

The dataset is current-voltage(I-V) features data obtained by I-V features extracted algorithm for the brand A modules under damp heat indoor accelerated test which is up to 3000 hours. The measurement were taken every 500h. The raw data is provided by SunEdison company. The I-V features include max power(Pmp), short circuit current(Isc), current at max power(Imp), fill factor(FF), series resistance(Rs), shunt resistance(Rsh), open circuit voltage(Voc), voltage at max power(Vmp). Rsh is too noisy to contain for modeling. After checking the correlation between Isc, Imp, Voc, Vmp, FF, Rs. FF, Rs, Vmp are highly correlated, so we randomly select one to be

contained in the model. Here we choose I_{sc} , I_{mp} , R_s and V_{oc} to be contained in the model and these four I-V features show no indication of high correlation. The trend of the I-V features are related with the mechanisms of PV degradation. The variable 'dy' is time that has been converted into decimal year in which 1 means 1 year. We would use this dataset to build <SIMIR> model with time as stressor, four I-V features as mechanisms and max power as response.

Usage

```
data(IVfeature)
```

Format

A data frame with 20 rows and 9 variables:

Pmp max power at the measurement step

dy the exposure time after been converted into decimal year

Isc short circuit current

Imp current at max power

Rs series resistance

Voc open circuit voltage

Author(s)

Jiqi Liu, Alan Curran, Justin S. Fada, Xuan Ma, Wei-Heng Huang, Jennifer L. Braid, Roger H. French

Source

Solar Durability and Lifetime Extension (SDLE) Research Center, Case Western Reserve University

netSEM

Network Structure Equation Modeling

Description

The R package 'netSEM' conducts a net-SEM statistical analysis (network structural equation modeling) on a dataframe of coincident observations of multiple continuous variables.

References

1. Bruckman, Laura S., Nicholas R. Wheeler, Junheng Ma, Ethan Wang, Carl K. Wang, Ivan Chou, Jiayang Sun, and Roger H. French. "Statistical and Domain Analytics Applied to PV Module Lifetime and Degradation Science." *IEEE Access* 1 (2013): 384-403. doi:10.1109/ACCESS.2013.2267611
2. Bruckman, Laura S., Nicholas R. Wheeler, Ian V. Kidd, Jiayang Sun, and Roger H. French. "Photovoltaic Lifetime and Degradation Science Statistical Pathway Development: Acrylic Degradation." In *SPIE Solar Energy+ Technology*, 8825:88250D-8. International Society for Optics and Photonics, 2013. doi:10.1117/12.2024717

netSEMm *network Structural Equation Modeling (netSEM)*

Description

This function carries out netSEM

Usage

```
netSEMm(x, exogenous = NULL, endogenous = NULL, nlsInits = data.frame(a1 =
  1, a2 = 1, a3 = 1), str = FALSE)
```

Arguments

x	A dataframe. By default it considers all columns as exogenous variables, except first column which stores the system endogenous variable.
exogenous,	by default it considers all columns as exogenous variables except column number 1, which is the main endogenous response.
endogenous	A character string of the column name of the main endogenous OR a numeric number indexing the column of the main endogenous.
nlsInits	a data frame of initial vectors for nls. Each column corresponds to a coefficient. The data frame can be generated by the genInit() function. Each row is one initial vector. Currently the only nls function included is $y = a + b * \exp(c * x)$.
str	A boolean, whether or not this is a 'strength' type problem

Details

netSEM builds a network model of multiple continuous variables. Each pair of variables is tested for sensible pairing relation chosen from 7 pre-selected common functional forms in linear regression settings. Adjusted R-squared is used for model selection for every pair.

P-values reported in the "res.print" field of the return list contains the P-values of estimators of linear regression coefficients. The P-values are ordered in the common order of coefficients, i.e. in the order of increasing exponents. For example, in the quadratic functional form $y \sim b_0 + b_1x + b_2x^2$, the three P-values correspond to those of \hat{b}_0 , \hat{b}_1 and \hat{b}_2 , respectively. If there are less than 3 coefficients to estimate, the extra P-value field is filled with NA's.

Value

An object of class netSEM, which is a list of the following items:

- "table": A matrix. For each row, first column is the endogenous variable, second column is the predictor, the other columns show corresponding summary information: Best functional form, R-squared, adj-R-squared, P-value1, P-value2 and P-value3. The P-values correspond to those of estimators of linear regression coefficients. See details.

- "bestModels": A matrix. First dimension indicates predictors. The second dimension indicates endogenous variables. The i-j-th cell of the matrix stores the name of the best functional form corresponding to the j-th endogenous variable regressed on the i-th predictor.
- "allModels": A three dimensional list. The first dimension indicates predictors. The second dimension indicates endogenous variables. Third dimension indicates the fitting results of all 6 functional forms. The i-j-k-th cell of the list stores a "lm" object, corresponding to the j-th endogenous, i-th predictor and the k-th functional form.

The object has two added attributes:

- "attr(res.best, "Step")": A vector. For each variable, it shows in which step it is chosen to be significantly related to endogenous variable.
- "attr(res.best, "diag.Step")": A matrix. First dimension is predictors; second dimension is endogenous variables. Each cell shows in which step the pairwise relation is being fitted.

Examples

```
## Load the sample acrylic data set
data(acrylic)

## Run netSEM
ans <- netSEMM(acrylic)

## Subset dataset
res <- subsetData(ans,cutoff=c(0.3,0.6,0.8))

## Plot the network model with adjusted-R-squared of c(0.3,0.6,0.8)
plot(ans,res)

## Summary
summary(ans)

## Extract relations between IrradTot and IAD2
cf <- path(ans,from ="IAD2",to="IrradTot")
print(cf)

## Print three components of the result
ans$table
ans$bestModels
ans$allModels
```

path

Extract Path Model Equation

Description

Extract and display an equation of a pairwise path between two variables.

Usage

```
path(x, from, to, round = 3)
```

Arguments

x	object of class "netSEM", which is the return value of function netSEMm.
from	character string. Name of the predictor.
to	character string. Name of the endogenous variable.
round	a positive integer. The coefficients are rounded to this decimal place.

Details

Extract the "best" model between any two variables. The model name and the model equation are printed on screen. The model coefficients, as well as the model R object are also returned.

Value

a list of the following items:

- "model": the best fitted model.
- "model.print": a character string of the model equation.

See Also

[netSEMm](#)

Examples

```
##' ## Load the sample acrylic data set
data(acrylic)

## Run netSEM principle three
ans <- netSEMm(acrylic)

## Extract relations between IrradTot and IAD2
cf <- path(ans, from = "IAD2", to = "IrradTot")
print(cf)
```

pathwayPredict *Summary of predicted and observed response values along a pathway*

Description

Predict response variable values along a pathway

Usage

```
pathwayPredict(x, path, newdata = NULL)
```

Arguments

x An object of class "netSEM", the returned list from netSEMM.

path A string form for a pathway, the default output format from [pathwayRMSE](#).

newdata A data frame of the stress variable. The default is NULL.

Value

An object of class pathway, which is a list of the following items:

- "pathway": A string form that shows the pathway.
- "RMSE": A value of the root mean squared error.
- "Resp": A matrix. The first column is the observed response values and the second is the predicted response values.

See Also

[pathwayRMSE](#), [netSEMM](#)

Examples

```
# Load the sample acrylic data set
data(acrylic)
ans <- netSEMM(acrylic)
paths <- pathwayRMSE(ans,maxlen=3)
response <- pathwayPredict(ans, paths[10,2])
response
```

pathwayRMSE

Summary of pathway RMSE

Description

summarize root mean square error (RMSE) for direct and indirect pathway from netSEM result

Usage

```
pathwayRMSE(x, maxlen = 2, ...)
```

Arguments

x	An object of class "netSEM", the returned list from netSEMm().
maxlen	The maximum length of chosen mechanism.
...	A S3 generic/method consistency.

Details

pathwayRMSE gives a summary about RMSE

Value

A data frame of result. A summary of RMSE results for all of pathways.

See Also

[netSEMm](#), [pathwayPredict](#)

Examples

```
data(acrylic)
ans <- netSEMm(acrylic)
pathwayRMSE(ans,maxlen=2)
```

pet *A data frame of an degradation experiment of poly(ethylene-terephthalate) films*

Description

The data set is a study of photolysis and hydrolysis of poly(ethylene-terephthalate) films that contain an ultraviolet stabilizer additive. In this work, polymeric samples were exposed to UV light and moisture according to ASTM G-154 Cycle 4 standard accelerated weathering conditions. Resulting optical chemical changes were determined through optical and infrared (IR) spectroscopy. **Time** is the main exogenous variable and **YI** (yellowness index) is the endogenous variable (response). The other columns in the data set (**AbsEdge**, **UV.Stab.Bleaching**, **Crystallization**, and **ChainScission**) are values extracted from optical and IR absorbance spectra as single metrics and used as intermediate unit level endogenous (response) variables in the netSEM analysis.

Usage

```
data(pet)
```

Format

A 37 by 6 data frame of continuous variables:

YI Yellowness index of PET film

Time Time exposed to ASTM G-154 Cycle 4 conditions

AbsEdge Fundamental Absorption Edge - Optical absorbance at 312 nm

UVStabBleaching Ultraviolet Stabilizer Bleaching - Optical absorbance at 340 nm

Crystallization IR spectroscopy peak of relative crystallinity - IR absorbance at 975 wavenumber

ChainScission IR carbonyl peak - IR absorbance at 1711 wavenumber

Author(s)

Devin A. Gordon

Source

Solar Durability and Lifetime Extension (SDLE) Research Center, Case Western Reserve University

plot.netSEM

Plotting of netSEM diagram

Description

Plot netSEM result plot.netSEM plots a structural equation network model diagram based on best functional form for each selected pairwise variable.

Usage

```
## S3 method for class 'netSEM'
plot(x, res, plot.save = FALSE, filename = NULL,
     style = TRUE, label = TRUE, ...)
```

Arguments

x	An object of class "netSEM", the returned list from netSEMm. Plotting uses the first element of this list (table) in which the first column of it is endogenous variable, second column is variable and other columns are corresponding best functional form, r-squared, adj-r-squared, P-value1, P-value2 and P-value3.
res	An object of class "subsetData", the returned list from subsetData. Stronger Solid lines represent relationship with higher adjusted R-sqr and weak dotted lines with less than the first cutoff.
plot.save	True/False, it saves the network diagram plot as png file. The default is false.
filename	A character string naming a file to save as a png file.
style	True/False, it plots the first interval in the network diagram with dotted weak line. The default is True.
label	True/False, it use label to express model and Adj-R2 in the path between variables. The default is True.
...	A S3 generic/method consistency.

Value

An html style plot of pairwise relationship pathway diagram between exogenous variables and an endogenous variable. Arrows show relationships between each variable with given statistical relations along the connection lines.

See Also

[netSEMm](#)

Examples

```
# Load acrylic data set
data(acrylic)
# Build a semi-gSEM model
ans <- netSEMM(acrylic)
# Subset dataset with three cutoff
res <- subsetData(ans,cutoff=c(0.3,0.6,0.8))
# Plot the network model
plot(ans,res)
# plot the network diagram and save as 'semplot.png' file
#plot(ans,res,plot.save=TRUE,filename=c("semplot"))
```

subsetData

Subset of the result table

Description

This function creat different Subset of the result table based on the cutoffs.

Usage

```
subsetData(x, cutoff = c(0.2, 0.5, 0.8))
```

Arguments

x	An object of class "netSEM", the returned list from netSEMM. Plotting uses the first element of this list (res.print) in which the first column of it is endogenous variable, second column is variable and other columns are corresponding best functional form, r-squared, adj-r-squared, P-value1, P-value2 and P-value3.
cutoff	A threshold value for adjusted R-squared. The maximum number of cutoff is 3.

Value

The dataframe of different subset of the result table.

Examples

```
# Load acrylic data set
data(acrylic)
# Build a semi-gSEM model
ans <- netSEMM(acrylic)
# Subset dataset with different cutoff
res <- subsetData(ans,cutoff=c(0.2))
res <- subsetData(ans,cutoff=c(0.2,0.5))
res <- subsetData(ans,cutoff=c(0.2,0.5,0.8))
```

summary.netSEM *Summary of net-SEM*

Description

summarize netSEM result

Usage

```
## S3 method for class 'netSEM'  
summary(object, ...)
```

Arguments

object	An object of class "netSEM", the returned list from netSEMm().
...	A S3 generic/method consistency.

Details

summary.netSEM gives a summary about the net-SEM analysis.

Value

NULL. A summary of data and fitting result is printed on screen.

Examples

```
data(acrylic)  
ans <- netSEMm(acrylic)  
summary(ans)
```


Index

*Topic **datasets**

- acrylic, [2](#)
- backsheet, [3](#)
- crack, [4](#)
- IVfeature, [6](#)
- pet, [13](#)

acrylic, [2](#)

backsheet, [3](#)

crack, [4](#)

findrelation, [5](#)

genInit, [6](#)

IVfeature, [6](#)

netSEM, [7](#)

netSEM-package (netSEM), [7](#)

netSEmm, [8](#), [10–12](#), [14](#)

path, [9](#)

pathwayPredict, [11](#), [12](#)

pathwayRMSE, [11](#), [12](#)

pet, [13](#)

plot.netSEM, [14](#)

subsetData, [15](#)

summary.netSEM, [16](#)