Package ‘composites’

September 30, 2020

Type Package

Title Compositional, Multivariate and Univariate Time Series Outlier Ensemble

Version 0.1.0

Maintainer Sevvandi Kandanaarachchi <sevvandik@gmail.com>

Description An ensemble of time series outlier detection methods that can be used for compositional, multivariate and univariate data. It uses the four R packages ‘forecast’, ‘tsoutliers’, ‘otsad’ and ‘anomalize’ to detect time series outliers.

License GPL-3

Encoding UTF-8

LazyData true

Imports otsad, tsoutliers, forecast, anomalize, dplyr, tidble, rlang, pracma, dobin, ICS, fastICA, gridExtra, grid, ggplot2, tidyr, kableExtra

RoxygenNote 7.1.1

Suggests knitr, rmarkdown, tourr, stringr, broom, rgdal

VignetteBuilder knitr

Depends R (>= 3.4.0)

URL https://sevvandi.github.io/composits/

NeedsCompilation no

Author Sevvandi Kandanaarachchi [aut, cre]
(https://orcid.org/0000-0002-0337-0395), Patricia Menendez [aut] (https://orcid.org/0000-0003-0701-6315), Ursula Laa [aut] (https://orcid.org/0000-0002-0249-6439), Ruben Loaiza-Maya [aut] (https://orcid.org/0000-0003-4132-8355)

Repository CRAN

Date/Publication 2020-09-30 09:50:02 UTC
animate_ts_ensemble

Show tour animation of the data points.

Description

Show tour animation of the data points.

Usage

```r
animate_ts_ensemble(
  obj = NULL,
  X = NULL,
  method = NULL,
  edges = NULL,
  max_frames = Inf
)
```

Arguments

- `obj`: The output from `comp_tsout_ens` or `mv_tsout_ens` functions.
- `X`: The data matrix used as input to `mv_tsout_ens` (not needed if `obj` is output from `comp_tsout_ens`).
- `method`: The dimension reduction method to apply before running the tour (if `NULL` tour on the full data space).
- `edges`: Set to "all" to connect points by time index, "outlying" to connect tagged outliers to previous and following points.
- `max_frames`: The maximum number of bases to generate in the grand tour (default is `Inf`).
Examples

set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)

animate_ts_ensemble(X = X, max_frames = 10)

out1 <- mv_tsout_ens(X, compr=2, fast=FALSE)
animate_ts_ensemble(out1, X, max_frames = 10)

X <- X/rowSums(X)
out2 <- comp_tsout_ens(X, ncomp = 3, compr=2, fast=FALSE)
animate_ts_ensemble(out2, method = "dobin", max_frames = 10)

---

Apportions outlier scores to composites.

**Description**

Apportions outlier scores to composites.

**Usage**

`apportion_scores_comp(obj)`

**Arguments**

- **obj**: The output of `comp_tsout_ens`

**Value**

A list with the following components:

- **scores_out**: The apportioned scores for outliers for timepoints in `mv_tsout_ens$outliers` or `comp_tsout_ens$outliers`.
- **scores_all**: The apportioned scores for outliers for timepoints in `mv_tsout_ens$all` or `comp_tsout_ens$all`.
Examples

```r
set.seed(1)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
X <- X/rowSums(X)
out <- comp_tsout_ens(X, compr=2, fast=FALSE)
apportioned <- apportion_scores_comp(out)
```

apportion_scores_mv  
Apportions outlier scores to composites.

Description

Apportions outlier scores to composites.

Usage

```r
apportion_scores_mv(obj)
```

Arguments

- `obj`  
The output of `comp_tsout_ens`

Value

A list with the following components:

- `scores_out`  
The apportioned outlier scores of selected outliers as per `codemv_tsout_ens`.
- `scores_all`  
The apportioned outlier scores of all identified outliers.

Examples

```r
set.seed(1)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
```
comp Tsout ens

Perform composite time series outlier ensembling.

Description

Performs composite time series outlier ensembling.

Usage

comp Tsout ens(
  x,
  m = NULL,
  ncomp = 2,
  sds = 1,
  rept = 1,
  compr = 2,
  rat = 0.05,
  fast = TRUE
)

Arguments

x A data frame or a matrix object containing a multivariate time series
m Variable indicating dimension reduction methods. Default is set to using all 4
ncomp The number of components for each dimension reduction method. Default is set
to 2.
sds The random seed for generating a no-outlier time series.
rept The number of repetitions for generating a no-outlier time series.
compr To adjust for multiple testing, the results of the ensemble are compared with the
results of a time series without outliers. If compr = 1, a time series is simulated
as in simulate_comp_ts without outliers. If compr = 2, the top outliers are
removed from the outlier series and interpolated values are used for those time
points. If compr = 3 both methods of simulation are used for comparison.
rat A comparison is done with the outliers removed time series. The variable rat
denotes the ratio of outliers to be removed as a proportion of the whole dataset
for this comparison.
fast For faster computation skip ICS decomposition method.

Example

x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
out <- mv Tsout ens(X, compr=2, fast=FALSE)
apportioned <- apportion_scores_mv(out)
Value

A list with the following components:

- **outliers**: The outliers detected from the multivariate ensemble after comparing with the comparison time series without outliers.
- **all**: All the outliers detected from the multivariate ensemble.
- **outmat**: A matrix with outlier scores organised by outlier method.
- **wts**: The weights of the outlier detection methods.
- **pca_loadings**: The basis vectors from PCA.
- **dobin_loadings**: The basis vectors from DOBIN. See R package `dobin` for more details.
- **ics_loadings**: The basis vectors from ICS. See R package `ICS` for more details.
- **ica_loadings**: The basis vectors from Independent Component Analysis.
- **decomp_wts**: Each decomposition method has several components. For example if `ncomp=2`, then there are 2 PC components, 2 DOBIN components, etc ... The weight of each component is given different and depends on the decomposition method. These weights are given in `decomp_wts`.
- **outmat4D**: A 4D array with outlier scores organised by outlier method, decomposition method, components for each decomposition method and time.
- **comp_loadings**: The unconstrained basis vectors on the simplex.
- **comp_coords**: The unconstrained coordinates of the composite time series data.

Examples

```r
set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
X <- X/rowSums(X)
out <- comp_tsout_ens(X, compr=2, fast=FALSE)
```

**draw_table**

Draws table from `comp_tsout_ens` or `mv_tsout_ens` output using `tableGrob`.

**Description**

Draws table from `comp_tsout_ens` or `mv_tsout_ens` output using `tableGrob`.
**Usage**

draw_table(obj, uniq_dates = NULL)

**Arguments**

obj       The output from comp_tsout_ens or mv_tsout_ens functions.
uniq_dates An optional parameter to pass in the dates for the dataset.

**Value**

Draws a table using R packages grid and gridExtra.

**Examples**

```r
set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
out <- mv_tsout_ens(X, compr=2, fast=FALSE)
draw_table(out)
```

**Description**

Draws an html table from comp_tsout_ens or mv_tsout_ens output using kableExtra.

**Usage**

draw_table_html(obj, uniq_dates = NULL)

**Arguments**

obj       The output from comp_tsout_ens or mv_tsout_ens functions.
uniq_dates An optional parameter to pass in the dates for the dataset.
get_coords

Computes unconstrained null space coordinates for compositional data.

Usage

get_coords(x)

Arguments

x  Compositional data in a dataframe or matrix. The rows need to add up to a constraint value

Value

A list with the following components:

y  The unconstrained coordinates of the input
vec  The basis vectors for the null space coordinates

Examples

set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
out <- mv_tsout_ens(X, compr=2, fast=FALSE)
draw_table_html(out)
mv_tsout_ens

X <- X/apply(X, 1, sum)
out <- get_coords(X)

mv_tsout_ens

Performs multivariate time series outlier ensembling.

Description

Performs multivariate time series outlier ensembling.

Usage

mv_tsout_ens(
  x,
  m1 = NULL,
  ncomp = 2,
  sds = 1,
  rept = 1,
  compr = 2,
  rat = 0.05,
  fast = TRUE
)

Arguments

x A data frame or a matrix object containing a multivariate time series
m1 Variable indicating dimension reduction methods. Default is set to using all 4 methods: PCA, DOBIN, ICS and ICA.
ncomp The number of components for each dimension reduction method. Default is set to 2.
sds The random seed for generating a no-outlier time series.
rept The number of repetitions for generating a no-outlier time series.
compr To adjust for multiple testing, the results of the ensemble are compared with the results of a time series without outliers. If compr =1, a time series is simulated as in simulate_comp_ts without outliers. If compr = 2, the top outliers are removed from the outlier series and interpolated values are used for those time points. If compr = 3 both methods of simulation are used for comparison.
rat A comparison is done with the outliers removed time series. The variable rat denotes the ratio of outliers to be removed as a proportion of the whole dataset for this comparison.
fast For faster computation skip ICS decomposition method.
Value

A list with the following components:

- **outliers**: The outliers detected from the multivariate ensemble after comparing with the comparison time series without outliers.
- **all**: All the outliers detected from the multivariate ensemble.
- **outmat**: A matrix with outlier scores organised by outlier method.
- **wts**: The weights of the outlier detection methods.
- **pca_loadings**: The basis vectors from PCA.
- **dobin_loadings**: The basis vectors from DOBIN. See R package dobin for more details.
- **ics_loadings**: The basis vectors from ICS. See R package ICS for more details.
- **ica_loadings**: The basis vectors from Independent Component Analysis.
- **decomp_wts**: Each decomposition method has several components. For example if ncomp=2, then there are 2 PC components, 2 DOBIN components, etc... The weight of each component is given different and depends on the decomposition method. These weights are given in decomp_wts.
- **outmat4D**: A 4D array with outlier scores organised by outlier method, decomposition method, components for each decomposition method and time.

Examples

```r
set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
dx2 <- sample(1:100, n, replace=TRUE)
dx3 <- sample(1:100, n, replace=TRUE)
dx4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
out <- mv_tsout_ens(X, m1=c(1,2,4), compr=2)
```

Description

The final graph shows the data points projected onto the first two components, together with the loadings as axes.
plot_decomposed

Usage

plot_biplot(obj, X = NULL, method = "pca", edges = NULL)

Arguments

obj The output from comp_tsout_ens or mv_tsout_ens functions.
X The data matrix used as input to mv_tsout_ens (not needed if obj is output from comp_tsout_ens).
method The decomposition method, choose between "pca" (default), "dobin", "ics" or "ica".
edges Set to "all" to connect points by time index, "outlying" to connect tagged outliers to previous and following points.

Value

A ggplot showing the biplot.

Examples

set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)

out1 <- mv_tsout_ens(X, compr=2, fast=FALSE)
plot_biplot(out1, X = X, method = "pca")

X <- X/rowSums(X)
out2 <- comp_tsout_ens(X, compr=2, fast=FALSE)
plot_biplot(out2)

plot_decomposed

Plot decomposed time series from comp_tsout_ens or mv_tsout_ens output.

Description

Plot decomposed time series from comp_tsout_ens or mv_tsout_ens output.
plot_decomposed_all

Usage

plot_decomposed_all(obj, X = NULL)

Description

Plot all decomposed time series from comp_tsout_ens or mv_tsout_ens output.

Usage

plot_decomposed_all(obj, X = NULL)
Simulations

Arguments

- **obj**
  The output from `comp_tsout_ens` or `mv_tsout_ens` functions.
- **X**
  The data matrix used as input to `mv_tsout_ens` (not needed if `obj` is output from `comp_tsout_ens`).

Value

A ggplot showing the time series with facets by decomposition method.

Examples

```r
set.seed(100)
n <- 600
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 300
x2 <- sample(1:100, n, replace=TRUE)
x3 <- sample(1:100, n, replace=TRUE)
x4 <- sample(1:100, n, replace=TRUE)
X <- cbind.data.frame(x, x2, x3, x4)
out <- mv_tsout_ens(X, m1=c(1,2,4), compr=2)
plot_decomposed_all(out, X=X)
```

Simulations

*Function to simulate compositional time series data*

Description

Function to simulate compositional time series data

Usage

```r
Simulations(N, TT, K, A, B, C, mu, D, outliers_discre, q)
```

Arguments

- **N**
  The number of categories in the composition
- **TT**
  The time series length
- **K**
  The state vector dimension
- **A**
  The N x K matrix of factor loadings in the observation equation
- **B**
  The K x K autoregressive matrix of the transition equation
- **C**
  The K x K matrix determining the magnitude of the persistent outliers
- **mu**
  The K-dimensional intercept vector in the transition equation
- **D**
  A K x K matrix determining the variance-covariance matrix of the error term
outliers_discre
An R x 3 matrix of discretionary outliers. R denotes the number of discretionary outliers. The first, second and third columns denote the time position, the composite position and the magnitude of the outliers.

q
Probability of persistent outlier eventuating

Value
A list with the following components:

datasim  A TT x K data frame with the generated time series compositional data.
outliers_persist  A matrix indicating the time location of the persistent outliers (first column) and the factors (or states) where the outlier eventuates (second column).
outliers_discre  A matrix equivalent to the function argument provided by the user.
outliers_timeloc  A vector with the time location of all the outliers.

Examples

set.seed(2000)
N <- 30
K <- 2
TT <- 500
A <- matrix(rnorm(N*K, 0, 0.3), N, K)
B <- matrix(c(0.8,0,0,0.5), K, K)
C <- matrix(c(5,0,0,4), K, K)
mu <- c(0.3, 0.7)
D <- matrix(c(0.4,0,0,0.4), K, K)
outliers_discre <- matrix(c(117, 2, 10, 40, 8, 200), 2, 3, byrow = TRUE)
q <- 0.005
y <- Simulations(N = N,
TT = TT,
K = K,
A = A,
B = B,
C = C,
mu = mu,
D = D,
outliers_discre = outliers_discre,
q = q)
spanish_morte

A dataset containing mortality counts in Spain by state.

Description

This dataset contains mortality counts in Spain by state from the 12th of July 2018 until the 29th of July 2020.

Usage

spanish_morte

Format

A list of two items.
1. The compositional dataset of mortality proportions by state.
2. The relevant dates.

uv_tsout_ens

Performs univariate time series outlier ensemble.

Description

Performs univariate time series outlier ensemble.

Usage

uv_tsout_ens(x, frequency = 1, dates = NULL)

Arguments

x A univariate time series as either a ts object or a vector.
frequency The frequency associated with the time series
dates The dates associated with the time series. This is needed for the package anomalize. If not explicitly set dates are set at a frequency 1 ending on the system date.

Value

A list with the following components:

outliers The outliers detected, repeated if detected by multiple outlier methods.
forecastOut The outliers detected R package forecast.
stsoutliersOut The outliers detected R package tsoutliers.
otsadOut The outliers detected R package otsad.
anomalizeOut The outliers detected R package anomalize.
outmat A matrix containing zeros and ones, with ones representing time points identified as outliers from different methods.
Examples

```r
set.seed(100)
n <- 500
x <- sample(1:100, n, replace=TRUE)
x[25] <- 200
x[320] <- 270
df <- data.frame(timestamp=1:n, value=x)
plot(ts(df$value))
oout <- uv_tsout_ens(x)
oout
```
Index

* datasets
  - spanish_morte, 15

animate_ts_ensemble, 2
apportion_scores_comp, 3
apportion_scores_mv, 4
comp_tsout_ens, 5
draw_table, 6
draw_table_html, 7
get_coords, 8
mv_tsout_ens, 9
plot_biplot, 10
plot_decomposed, 11
plot_decomposed_all, 12

Simulations, 13
spanish_morte, 15
uv_tsout_ens, 15