

Package: tsdataleaks (via r-universe)

October 16, 2024

Type Package

Title Exploit Data Leakages in Time Series Forecasting Competitions

Version 2.1.1

Description Forecasting competitions are of increasing importance as a mean to learn best practices and gain knowledge. Data leakage is one of the most common issues that can often be found in competitions. Data leaks can happen when the training data contains information about the test data. For example: randomly chosen blocks of time series are concatenated to form a new time series, scale-shifts, repeating patterns in time series, white noise is added in the original time series to form a new time series, etc. 'tsdataleaks' package can be used to detect data leakages in a collection of time series.

License GPL (>= 2)

URL <https://github.com/thiyangt/tsdataleaks>

BugReports <https://github.com/thiyangt/tsdataleaks/issues>

Depends R (>= 3.6.0)

Imports stats, tibble (>= 1.4.1), ggplot2 (>= 3.0.0), dplyr (>= 1.0.0), tidyr (>= 1.1.0), slider, purrr, utils, cowplot, plyr, viridis

Encoding UTF-8

RoxygenNote 7.2.3

Suggests testthat (>= 2.1.0), knitr, rmarkdown

VignetteBuilder knitr

Repository <https://thiyangt.r-universe.dev>

RemoteUrl <https://github.com/thiyangt/tsdataleaks>

RemoteRef HEAD

RemoteSha 697d606273af3e3472e0b62ebc80f2589c10de92

Contents

find_dataleaks	2
reason_dataleaks	3
ts.match	4
viz_dataleaks	5
Index	6

find_dataleaks	<i>Correlation calculation based on rolling window with overlapping observations.</i>
----------------	---

Description

Correlation calculation based on rolling window with overlapping observations.

Usage

```
find_dataleaks(lstx, h, cutoff = 1)
```

Arguments

lstx	list of time series
h	length of forecast horizon
cutoff	benchmark value for corr absolute value, default 1

Value

list of matching quantities

Examples

```
a = rnorm(15)
lst <- list(
  a = a,
  b = c(a[10:15], rnorm(10), a[1:5], a[1:5]),
  c = c(rnorm(10), a[1:5])
)
find_dataleaks(lst, h=5)
#' a = rnorm(15)
lst <- list(
  x= a,
  y= c(rnorm(10), a[1:5])
)

find_dataleaks(lst, h=5)

# List without naming elements
lst <- list(
```

```

a,
c(rnorm(10), a[1:5], a[1:5]),
rnorm(10)
)
find_dataleaks(lst, h=5)

```

reason_dataleaks	<i>Correlation calculation based on rolling window with overlapping observations.</i>
------------------	---

Description

Correlation calculation based on rolling window with overlapping observations.

Correlation calculation based on rolling window with overlapping observations.

Usage

```
reason_dataleaks(lstx, finddataleaksout, h, ang = 0)
```

```
reason_dataleaks(lstx, finddataleaksout, h, ang = 0)
```

Arguments

lstx	list of time series
finddataleaksout	list, the output generated from find_dataleaks function
h	length of the window size
ang	angle at which the tick and axis labels should be displayed (default 0)

Value

matrix visualizing the output

matrix visualizing the output

Examples

```

a = rnorm(15)
lst <- list(
  a = a,
  b = c(a[10:15], rnorm(10), a[1:5]+10, a[1:5]),
  c = c(rnorm(10), a[1:5])
)
f1 <- find_dataleaks(lst, h=5)
reason_dataleaks(lst, f1, h=5)

# List without naming elements
lst <- list(
  a,

```

```

    c(rnorm(10), a[1:5], a[1:5]),
    rnorm(10)
  )
  f2 <- find_dataleaks(lst, h=5)
  reason_dataleaks(lst, f2, h=5)
  a = rnorm(15)
  lst <- list(
    a = a,
    b = c(a[10:15], rnorm(10), a[1:5], a[1:5]),
    c = c(rnorm(10), a[1:5])
  )
  f1 <- find_dataleaks(lst, h=5)
  reason_dataleaks(lst, f1, h=5)

```

ts.match	<i>Correlation calculation based on rolling window with overlapping observations.</i>
----------	---

Description

Correlation calculation based on rolling window with overlapping observations.

Usage

```
ts.match(x, y, cutoff = 1)
```

Arguments

x	time series
y	subsection of the time series to map
cutoff	benchmark value for corr, default 1

Value

Pearson's correlation coefficient between x and y

Examples

```

x <- rnorm(15)
y <- -x[6:10]
x <- c(x, y)
ts.match(x, y, 1)
z <- rnorm(5)
ts.match(x, z)

```

viz_dataleaks	<i>Correlation calculation based on rolling window with overlapping observations.</i>
---------------	---

Description

Correlation calculation based on rolling window with overlapping observations.

Usage

```
viz_dataleaks(finndataleaksout)
```

Arguments

```
finndataleaksout  
list, the output generated from find_dataleaks function
```

Value

matrix visualizing the output

Examples

```
a = rnorm(15)  
lst <- list(  
  a = a,  
  b = c(a[10:15]+rep(8,6), rnorm(10), a[1:5], a[1:5]),  
  c = c(rnorm(10), a[1:5]),  
  d = rnorm(10)  
)  
f1 <- find_dataleaks(lst, h=5)  
viz_dataleaks(f1)
```

```
a = rnorm(15)  
lst <- list(  
  x= a,  
  y= c(rnorm(10), a[1:5])  
)
```

```
f2 <- find_dataleaks(lst, h=5)  
viz_dataleaks(f2)
```

```
# List without naming elements  
lst <- list(  
  a,  
  c(rnorm(10), a[1:5], a[1:5]),  
  rnorm(10)  
)  
f3 <- find_dataleaks(lst, h=5)  
viz_dataleaks(f3)
```

Index

`find_dataleaks`, 2
`reason_dataleaks`, 3
`ts.match`, 4
`viz_dataleaks`, 5