

Package ‘REDI’

June 7, 2023

Title Robust Exponential Decreasing Index

Version 1.0.0

Maintainer Alexia Grenouillat <alexia.grenouillat00@gmail.com>

Description Implementation of the Robust Exponential Decreasing Index (REDI), proposed in the article by Issa Moussa, Arthur Leroy et al. (2019) <<https://bmjopensem.bmj.com/content/bmjosem/5/1/e000573.full.pdf>>. The REDI represents a measure of cumulated workload, robust to missing data, providing control of the decreasing influence of workload over time. Various functions are provided to format data, compute REDI, and visualise results in a simple and convenient way.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports dplyr, ggplot2, tidyr, tibble, magrittr, rlang, lubridate

NeedsCompilation no

Author Alexia Grenouillat [aut, cre],
Arthur Leroy [aut]

Repository CRAN

Date/Publication 2023-06-07 13:10:02 UTC

R topics documented:

compute_redi	2
format_data	2
loop_redi	3
plot_redi	4
redi	5
simu_db	6

Index	8
--------------	----------

compute_redi	<i>Compute REDI for a specific input</i>
--------------	--

Description

Compute REDI for a specific input

Usage

```
compute_redi(data, coef = 0.1)
```

Arguments

data	A tibble or data frame, containing an Input column (with the Date format) and an Output column. A simple vector of workload values, pre-sorted by chronological order, can also be provided.
coef	A number corresponding to the lambda coefficient, controlling the decay of the exponential weights.

Value

A number, corresponding to the REDI value at the last Input time, computed over the whole period.

Examples

```
data <- simu_db()
compute_redi(data = data, coef = 0.1)
```

format_data	<i>Format the dataset to the syntax of REDI functions</i>
-------------	---

Description

Format the dataset to the syntax of REDI functions

Usage

```
format_data(
  data,
  input = 1,
  output = 2,
  by = "day",
  format = "%Y%m%d",
  summary_duplicate = mean
)
```

Arguments

data	A tibble or data frame containing one column indicating time and another indicating the quantity for which we want to compute REDI.
input	A character or a number, indicating whether the name or the index of the input column (time).
output	A character or a number, indicating whether the name or the index of the output column (workload).
by	A number or a character string, indicating the reference time period between two observations. Possible values are 'day', 'week', 'month', 'year', or any arbitrary number. See documentation of the 'seq()' for additional information if necessary. Default is 'day'.
format	A character string, indicating the date format of the input. Please read <code>lubridate::as_date()</code> . Default is '%Y%m%d'.
summary_duplicate	A function, used to summarise Output values for duplicated Input values. Default is mean.

Value

A tibble with Input and Output columns and explicit missing values between observations.

Examples

```
TRUE
```

loop_redi	<i>Compute the evolution of REDI over successive inputs</i>
-----------	---

Description

Compute the evolution of REDI over successive inputs

Usage

```
loop_redi(data, coef = 0.1)
```

Arguments

data	A tibble or data frame, containing an Input column (with the Date format) and an Output column. A simple vector of workload values, pre-sorted by chronological order can also be provided.
coef	A number corresponding to the lambda coefficient, controlling the decay of the exponential weights. Default is 0.1.

Value

A tibble similar to data, containing an additional REDI column computed over the successive input values.

Examples

```
data <- simu_db()
loop_redi(data = data, coef = 0.1)
```

plot_redi *Display the evolution of REDI over time and data points.*

Description

Display the evolution of REDI over time and data points.

Usage

```
plot_redi(
  redi,
  plot_data = TRUE,
  x_axis = "Input",
  y_axis = "Output",
  alpha = 0.2,
  size = 1
)
```

Arguments

redi	A tibble or data frame containing 4 mandatory columns : Input, Output, REDI and Lambda. One can use the redi() function to get this tibble under the right format.
plot_data	A boolean, indicating whether original data should be displayed. Default is TRUE.
x_axis	A character string, label of the x-axis. Default is 'Input'.
y_axis	A character string, label of the y-axis. Default is 'Output'.
alpha	A number, between 0 and 1, controlling the transparency of data points. Default is 0.2.
size	A number, controlling the size of the data points. Default is 1.

Value

Graph of the evolution of REDI over time, possibly for different values of Lambda, along with the original data points.

Examples

```
TRUE
```

redi *Compute REDI for all observed and missing input values in a dataset*

Description

Wrapper function that converts the dataset to the adequate format, compute values of REDI for each Input values, display a generic plot of the results and return a tibble containing both data and corresponding REDI values.

Usage

```
redi(
  data,
  coef = c(0.05, 0.1, 0.5),
  input = 1,
  output = 2,
  plot = TRUE,
  by = "day",
  format = "%Y%m%d",
  summary_duplicate = mean
)
```

Arguments

data	A tibble or a data frame, containing an Input column that is used as reference for the observations (e.g. time for longitudinal data), and an Output column specifying the observed values of interest (the workload).
coef	A number or vector, containing the values of the lambda coefficient used in the REDI computations, controlling the decay of the exponential weights. Default is <code>c(0.05, 0.1, 0.5)</code> .
input	A character or a number, indicating the name or the index of the Input column (time).
output	A character or a number, indicating the name or the index of the Output column (workload).
plot	A boolean, indicating whether results should be displayed. is TRUE.
by	A number or a character string, indicating the reference time period between two observations. Possible values are 'day', 'week', 'month', 'year', or any arbitrary number. See documentation of the 'seq()' for additional information if necessary. Default is 'day'.
format	A character string, indicating the date format of the input. Please read <code>lubridate::as_date()</code> . Default is '%Y%m%d'.
summary_duplicate	A function, used to summarise Output values for duplicated Input values. Default is mean.

Value

A tibble containing 4 columns : Input (without duplicates), Output, Lambda and REDI, which corresponds to the vector returned by the `loop_REDI()` function.

Examples

```
data <- simu_db()
redi <- redi(data)
```

 simu_db

Generate a synthetic dataset tailored for REDI computations

Description

Simulate a complete training dataset, which may be representative of various applications. Several flexible arguments allow adjustment of the range of observed days, the distribution and the mean of Output values, as well as the ratio of missing data.

Usage

```
simu_db(
  start_date = "2022-01-01",
  end_date = "2023-01-01",
  by = "day",
  output_distrib = "Gaussian",
  ratio_missing = 0.5,
  mean = 50,
  var = 10,
  range_unif = c(0, 100)
)
```

Arguments

<code>start_date</code>	A date, indicating the starting time of observations. Default is '2022-01-01'.
<code>end_date</code>	A date, indicating the ending time of observations. Default is '2023-01-01'.
<code>by</code>	A number or a character string, indicating the reference time period between two observations. Possible values are 'day', 'week', 'month', 'year', or any arbitrary number. See documentation of the 'seq()' for additional information if necessary. Default is 'day'.
<code>output_distrib</code>	A character string, indicating the distribution of Output values. Possible values: 'Gaussian' (default), 'Uniform'.
<code>ratio_missing</code>	A number, between 0 and 1, indicating the ratio of missing values in the dataset. Default is 0.5.
<code>mean</code>	A number, indicating the mean value of the Gaussian distribution. Default is 50.
<code>var</code>	A number, indicating the variance of the Gaussian distribution. Default is 10.
<code>range_unif</code>	A vector, indicating the range of values for the Uniform distribution. Default is c(0,100).

Value

A full dataset of synthetic data.

Examples

```
## Generate a dataset with Gaussian measurements  
data = simu_db(output_distrib = 'Gaussian')
```

```
## Generate a dataset with Uniform measurements and 30% of missing data.  
data = simu_db(output_distrib = 'Uniform', ratio_missing = 0.3)
```

Index

`compute_redi`, 2

`format_data`, 2

`loop_redi`, 3

`plot_redi`, 4

`redi`, 5

`simu_db`, 6