Package 'WaveletMLbestFL'

January 20, 2025

Type Package

Title The Best Wavelet Filter-Level for Prepared Wavelet-Based Models

Version 0.1.0

Author Mr. Sandip Garai [aut, cre],
Dr. Ranjit Kumar Paul [aut]

Maintainer Mr. Sandip Garai <sandipnicksandy@gmail.com>

Description Four filters have been chosen namely 'haar', 'c6', 'la8', and 'bl14' (Kindly refer to 'wavelets' in 'CRAN' repository for more supported filters). Levels of decomposition are 2, 3, 4, etc. up to maximum decomposition level which is ceiling value of logarithm of length of the series base 2. For each combination two models are run separately. Results are stored in 'input'. First five metrics are expected to be minimum and last three metrics are expected to be maximum for a model to be considered good. Firstly, every metric value (among first five) is searched in every columns and minimum values are denoted as 'MIN' and other values are denoted as 'NA'. Secondly, every metric (among last three) is searched in every columns and maximum values are denoted as 'MAX' and other values are denoted as 'NA'. 'output' contains the similar number of rows (which is 8) and columns (which is number filter-level combinations) as of 'input'. Values in 'output' are corresponding 'NA', 'MIN' or 'MAX'. Finally, the column containing minimum number of 'NA' values is denoted as the best ('FL'). In special case, if two columns having equal 'NA', it has been checked among these two columns which one is having least 'NA' in first five rows and has been inferred as the best. 'FL_metrics_values' are the corresponding metrics values. 'WARIGAANbest' is the data frame (dimension: 1*8) containing different metrics of the best filter-level combination. More details can be found in Garai and others (2023) <doi:10.13140/RG.2.2.11977.42087>.

License GPL-3

Encoding UTF-8

Imports WaveletML, CEEMDANML, DescribeDF

RoxygenNote 7.2.1

NeedsCompilation no

Repository CRAN

Date/Publication 2023-04-13 09:20:02 UTC

Contents

---

best_column	*Finds The Best Column In A Data Frame Containing Eight Metrics Values In The Columns*

Description

In statistical modeling, multiple models need to be compared based on certain critera. The method described here uses eight metrics from "AllMetrics" package. 'input_df' is the data frame (at least two columns for comparison) containing metrics values in different rows of a column (which denotes a particular model's performance). First five metrics are expected to be minimum and last three metrics are expected to be maximum for a model to be considered good. Firstly, every metric value (among first five) is searched in every columns and minimum values are denoted as 'MIN' and other values are denoted as 'NA'. Secondly, every metric (among last three) is searched in every columns and maximum values are denoted as 'MAX' and other values are denoted as 'NA'. 'output_df' contains the similar number of rows (which is 8) and columns (which is number of models to be compared) as of 'input_df'. Values in 'output_df' are corresponding 'NA', 'MIN' or 'MAX'. Finally, the column containing minimum number of 'NA' values is denoted as the best column. 'min_NA_col' gives the name of the best column (model). 'min_NA_values' are the corresponding metrics values. 'BestColumn_metrics' is the data frame (dimension: 1*8) containing different metrics of the best column (model). 'best_column_results' is the final result (a list) containing all of these output elements. In special case, if two columns having equal 'NA', it will be checked among these two column which one is having least 'NA' in first five rows and will be inferred as the best. More details about 'AllMetrics' can be found in Garai (2023) <doi:10.13140/RG.2.2.18688.30723>.

Usage

```
best_column(input_df)
```

Arguments

input_df	Data Frame Containing Eight Metrics Values In Different Columns

Value

- best_column_results: A list containing four data frames

References

- Paul, R. K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. Soft Computing, 25(20), 12857-12873.

- Paul, R. K., & Garai, S. (2022). Wavelets based artificial neural network technique for forecasting agricultural prices. Journal of the Indian Society for Probability and Statistics, 23(1), 47-61.

- Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEMDAN decomposition and machine intelligence. Intelligent Systems with Applications, 18, 200202.

- Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., Barman, S. & Manjunatha, B. (2023). An MRA Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. International Journal of Environment and Climate Change, 13(5), 137-150.

- Garai, S. (2023). Package 'AllMetrics' Type Package Title Calculating Multiple Performance Metrics of a Prediction Model Version 0.1.0, Repository: https://cran.r-project.org/web/packages/AllMetrics/index.html

Examples

```
set.seed(123) # for reproducibility
Y <- rnorm(50, 25, 5)
model_1 <-CEEMDANML::carigaan(Y, ratio = 0.8, n_lag = 4)
model_2 <- CEEMDANML::carigas(Y, ratio = 0.8, n_lag = 4)
model_1_metrics_values <- model_1$Accuracy[,2]
model_2_metrics_values <- model_2$Accuracy[,2]
combined_results <- data.frame(cbind(model_1_metrics_values, model_2_metrics_values))
colnames(combined_results) <- c("model_1", "model_2")
best_model <- best_column(input_df = combined_results)
```

---

warigaan_best	*Provides The Best Wavelet Filter-Level Combination For WARIGAAN Model*

Description

Four filters have been chosen namely 'haar', 'c6', 'la8', and 'bl14'. (Kindly refer to 'wavelets' in 'CRAN' repository for more supported filters). Levels of decomposition are 2, 3, 4, etc. up to maximum decomposition level which is ceiling value of logarithm of length of the series base 2. For each combination 'warigaan' are run (from 'WaveletML' package). Results are stored in 'WARIGAANbest_input_df'. First five metrics are expected to be minimum and last three metrics are expected to be maximum for a model to be considered good. Firstly, every metric value (among first five) is searched in every columns and minimum values are denoted as 'MIN' and other values are denoted as 'NA'. Secondly, every metric (among last three) is searched in every columns and maximum values are denoted as 'MAX' and other values are denoted as 'NA'. 'WARIGAANbest_output_df' contains the similar number of rows (which is 8) and columns

(which is number filter-level combinations) as of 'WARIGAANbest_input_df'. Values in 'WARI-GAANbest_output_df' are corresponding 'NA', 'MIN' or 'MAX'. Finally, the column containing minimum number of 'NA' values is denoted as the best ('WARIGAANbest_FL). In special case, if two columns having equal 'NA', it has been checked among these two columns which one is having least 'NA' in first five rows and has been inferred as the best. 'WARIGAANbest_FL_metrics_values' are the corresponding metrics values. 'WARIGAANbest' is the data frame (dimension: 1*8) containing different metrics of the best filter-level combination. More details about 'AllMetrics' can be found in Garai et al. (2023) <doi:10.13140/RG.2.2.11977.42087>.

Usage

```
warigaan_best(df, col = 2, f_l = list(f_vals = c('haar', 'c6', 'la8', 'bl14'),
                                      l_vals = 2:floor(log2(length(Y)))))
```

Arguments

df	Data Frame Containing various time series data except in the the 1st column
col	Mention the column number to be analysed except the 1st column
f_l	filter and level of decomposition should be chosen in this format

Value

- WARIGAANbest_input_df: Data frame containing metrics values for WARIGAAN model different filter-level combinations
- WARIGAANbest_output_df: Data frame containing 'NA', 'MIN' or 'MAX' values as described earlier
- WARIGAANbest_FL: Best filter-level combination
- WARIGAANbest_FL_metrics_values: Corresponding metrics values of WARIGAANbest_FL
- WARIGAANbest: WARIGAANbest_FL with all metrics values

References

- Aldrich, E. (2020). wavelets: Functions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses. Repository:https://cran.r-project.org/web/packages/wavelets/index.html.
- Paul, R. K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. Soft Computing, 25(20), 12857-12873.
- Paul, R. K., & Garai, S. (2022). Wavelets based artificial neural network technique for forecasting agricultural prices. Journal of the Indian Society for Probability and Statistics, 23(1), 47-61.
- Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEM-DAN decomposition and machine intelligence. Intelligent Systems with Applications, 18, 200202.
- Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., Barman, S. & Manjunatha, B. (2023). An MRA Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. International Journal of Environment and Climate Change, 13(5), 137-150.

• Garai, S. Paul, R. K., Yeasin, M.(2023). Package 'WaveletML' Type Package Title Wavelet
  Decomposition Based Hybrid Machine Learning Models, Repository: https://cran.r-project.org/web/packages/WaveletN

Examples

```
# example_data
# Set the seed for reproducibility
set.seed(123)
# Define the values of N and CV
N <- 25
CV <- "(5-10)"
mu <- 20
# define named list with f_vals and l_vals
params <- list(f_vals = 'c18',
               l_vals = 2:3)
# Define the number of data points to generate for each combination of N and CV
n_data <- 3
cv_range <- as.numeric(strsplit(gsub("[()]", "", CV), "-")[[1]])
cv_values <- seq(cv_range[1], cv_range[2])
cv_sample <- sample(cv_values, n_data, replace = TRUE)
sd <- mu * (cv_sample/100)
data <- replicate(n_data, {
  d <- round(abs(rnorm(N, mean = mu, sd = sd)))
  d[d == 0] <- 1
  d
})
colnames(data) <- paste0("N_", N, "_CV_", CV, "_", 1:n_data)
# Add a column for sequential numbers
data_new <- data.frame(cbind("sl no" = 1:N, data))
# example
warigaan_best_model <- warigaan_best(df = data_new, col = 2, f_l = params)
warigaan_best_model
```

---

warigas_best	*Provides The Best Wavelet Filter-Level Combination For WARIGAS Model*

Description

Four filters have been chosen namely 'haar', 'c6', 'la8', and 'bl14'. (Kindly refer to 'wavelets' in 'CRAN' repository for more supported filters). Levels of decomposition are 2, 3, 4, etc. up to maximum decomposition level which is ceiling value of logarithm of length of the series base 2. For each combination 'warigas' are run (from 'WaveletML' package). Results are stored in 'WARIGASbest_input_df'. First five metrics are expected to be minimum and last three metrics are expected to be maximum for a model to be considered good. Firstly, every metric value (among first five) is searched in every columns and minimum values are denoted as 'MIN' and other values are denoted as 'NA'. Secondly, every metric (among last three) is searched in every columns and maximum values are denoted as 'MAX' and other values are denoted as 'NA'. 'WARIGAS-best_output_df' contains the similar number of rows (which is 8) and columns (which is number

filter-level combinations) as of 'WARIGASbest_input_df'. Values in 'WARIGASbest_output_df' are corresponding 'NA', 'MIN' or 'MAX'. Finally, the column containing minimum number of 'NA' values is denoted as the best ('WARIGASbest_FL). In special case, if two columns having equal 'NA', it has been checked among these two columns which one is having least 'NA' in first five rows and has been inferred as the best. 'WARIGASbest_FL_metrics_values' are the corresponding metrics values. 'WARIGASbest' is the data frame (dimension: 1*8) containing different metrics of the best filter-level combination. More details about 'AllMetrics' can be found in Garai et al. (2023) <doi:10.13140/RG.2.2.11977.42087>.

Usage

```
warigas_best(df, col = 2, f_l = list(f_vals = c('haar', 'c6', 'la8', 'bl14'),
                                      l_vals = 2:floor(log2(length(Y)))))
```

Arguments

df	Data Frame Containing various time series data except in the the 1st column
col	Mention the column number to be analysed except the 1st column
f_l	filter and level of decomposition should be chosen in this format

Value

- WARIGASbest_input_df: Data frame containing metrics values for WARIGAS model different filter-level combinations
- WARIGASbest_output_df: Data frame containing 'NA', 'MIN' or 'MAX' values as described earlier
- WARIGASbest_FL: Best filter-level combination
- WARIGASbest_FL_metrics_values: Corresponding metrics values of WARIGASbest_FL
- WARIGASbest: WARIGASbest_FL with all metrics values

References

- Aldrich, E. (2020). wavelets: Functions for Computing Wavelet Filters, Wavelet Transforms and Multiresolution Analyses. Repository:https://cran.r-project.org/web/packages/wavelets/index.html.
- Paul, R. K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. Soft Computing, 25(20), 12857-12873.
- Paul, R. K., & Garai, S. (2022). Wavelets based artificial neural network technique for forecasting agricultural prices. Journal of the Indian Society for Probability and Statistics, 23(1), 47-61.
- Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEM-DAN decomposition and machine intelligence. Intelligent Systems with Applications, 18, 200202.
- Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., Barman, S. & Manjunatha, B. (2023). An MRA Based MLR Model for Forecasting Indian Annual Rainfall Using Large Scale Climate Indices. International Journal of Environment and Climate Change, 13(5), 137-150.

- Garai, S. Paul, R. K., Yeasin, M.(2023). Package 'WaveletML' Type Package Title Wavelet Decomposition Based Hybrid Machine Learning Models, Repository: https://cran.r-project.org/web/packages/WaveletM

Examples

```
# example_data
# Set the seed for reproducibility
set.seed(123)
# Define the values of N and CV
N <- 25
CV <- "(5-10)"
mu <- 20
# define named list with f_vals and l_vals
params <- list(f_vals = 'c18',
               l_vals = 2:3)
# Define the number of data points to generate for each combination of N and CV
n_data <- 3
cv_range <- as.numeric(strsplit(gsub("[()]", "", CV), "-")[[1]])
cv_values <- seq(cv_range[1], cv_range[2])
cv_sample <- sample(cv_values, n_data, replace = TRUE)
sd <- mu * (cv_sample/100)
data <- replicate(n_data, {
  d <- round(abs(rnorm(N, mean = mu, sd = sd)))
  d[d == 0] <- 1
  d
})
colnames(data) <- paste0("N_", N, "_CV_", CV, "_", 1:n_data)
# Add a column for sequential numbers
data_new <- data.frame(cbind("sl no" = 1:N, data))
# example
warigas_best_model <- warigas_best(df = data_new, col = 2, f_l = params)
warigas_best_model
```

Index