Package 'segen'

August 19, 2025

Type Package

Title Sequence Generalization Through Similarity Network

Version 2.0.0

Maintainer Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

Description Proposes an application for sequence prediction generalizing the similarity within the network of previous sequences.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.6)

Imports purrr (>= 0.3.4), ggplot2 (>= 3.3.5), readr (>= 2.1.2), lubridate (>= 1.7.10), imputeTS (>= 3.2), fANCOVA (>= 0.6-1), scales (>= 1.1.1), tictoc (>= 1.0.1), modeest (>= 2.4.0), moments (>= 0.14), greybox (>= 1.0.1), philentropy (>= 0.5.0), entropy (>= 1.3.1), Rfast (>= 2.0.6), narray (>= 0.4.1.1), fastDummies (>= 1.6.3), dtw (>= 1.23-1), digest (>= 0.6.31), furrr (>= 0.3.1), future (>= 1.33.0)

URL https://rpubs.com/giancarlo_vercellino/segen

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Giancarlo Vercellino [aut, cre, cph]

Repository CRAN

Date/Publication 2025-08-19 16:00:02 UTC

Contents

segen 2
time_features 4

Index **5**

segen *segen*

Description

Sequence Generalization Through Similarity Network

Usage

```
segen(
  df,
  seq_len = NULL,
  similarity = NULL,
  dist_method = NULL,
  rescale = NULL,
  smoother = FALSE,
  ci = 0.8,
  error_scale = "naive",
  error_benchmark = "naive",
  n_windows = 10,
  n_samp = 30,
  dates = NULL,
  seed = 42,
  use_parallel = FALSE,
  parallel_workers = NULL
)
```

Arguments

df	data.frame of time features (all numeric OR all categorical).
seq_len	integer, forecasting horizon. If NULL, auto-sampled.
similarity	numeric in (0,1), similarity quantile. If NULL, sampled.
dist_method	character. Options: "euclidean","manhattan","maximum","minkowski","correlation","dtw". If NULL, sampled from available methods (skips 'dtw' if pkg missing).
rescale	logical, rescale weights before normalization.
smoother	logical, apply loess smoothing for numeric features.
ci	numeric in (0,1), confidence level.
error_scale	"naive" or "deviation".
error_benchmark	"naive" or "average".
n_windows	integer, rolling validation windows.
n_samp	integer, random search samples.
dates	Date vector aligned with rows of df (optional).
seed	integer, RNG seed.

`use_parallel` logical, use furrr/future for parallel exploration.

`parallel_workers` NULL or integer, number of workers when parallel.

Value

list with exploration, history, best_model, time_log.

This function returns a list including:

- exploration: list of all not-null models, complete with predictions and error metrics
- history: a table with the sampled models, hyper-parameters, validation errors
- best_model: results for the best selected model according to the weighted average rank, including:
  - predictions: for continuous variables, min, max, q25, q50, q75, quantiles at selected ci, mean, sd, mode, skewness, kurtosis, IQR to range, risk ratio, upside probability and divergence for each point fo predicted sequences; for factor variables, min, max, q25, q50, q75, quantiles at selected ci, proportions, difformity (deviation of proportions normalized over the maximum possible deviation), entropy, upgrade probability and divergence for each point fo predicted sequences
  - testing_errors: testing errors for each time feature for the best selected model (for continuous variables: me, mae, mse, rmsse, mpe, mape, rmae, rrmse, rame, mase, smse, sce, gmrae; for factor variables: czekanowski, tanimoto, cosine, hassebrook, jaccard, dice, canberra, gower, lorentzian, clark)
  - plots: standard plots with confidence interval for each time feature
- time_log

Author(s)

Giancarlo Vercellino <giancarlo.vercellino@gmail.com>

Maintainer: Giancarlo Vercellino <giancarlo.vercellino@gmail.com> [copyright holder]

See Also

Useful links:

- https://rpubs.com/giancarlo_vercellino/segen

Examples

```
segen(time_features[, 1, drop = FALSE], seq_len = 30, similarity = 0.7, n_windows = 3, n_samp = 1)
```

`time_features` *time features example: IBM and Microsoft Close Prices*

Description

A data frame with with daily with daily prices for IBM and Microsoft since April 2020

Usage

```
time_features
```

Format

A data frame with 2 columns and 1324 rows.

Source

finance.yahoo.com

Index

∗ **datasets**
time_features, 4

segen, 2
segen-package *(*segen*)*, 2

time_features, 4