Package ‘proteus’

Type Package
Title Multiform Seq2Seq Model for Time-Feature Analysis
Version 1.0.0
Author Giancarlo Vercellino
Maintainer Giancarlo Vercellino <giancarlo.vercellino@gmail.com>
Description Seq2seq time-feature analysis based on variational model, with a wide range of distributions available for the latent variable.
License GPL-3
Encoding UTF-8
LazyData true
RoxygenNote 7.1.1
Depends R (>= 3.6)
Imports purrr (>= 0.3.4), abind (>= 1.4-5), ggplot2 (>= 3.3.3), readr (>= 1.4.0), stringr (>= 1.4.0), lubridate (>= 1.7.9.2), narray (>= 0.4.1), fANCOVA (>= 0.6-1), imputeTS (>= 3.1), modeest (>= 2.4.0), scales (>= 1.1.1), tictoc (>= 1.0.1), bizdays (>= 1.0.6), torch (>= 0.3.0), actuar (>= 3.1-1), VGAM (>= 1.1-5), moments (>= 0.14), dplyr (>= 1.0.2)
URL https://rpubs.com/giancarlo_vercellino/proteus
NeedsCompilation no
Repository CRAN
Date/Publication 2021-06-24 11:50:02 UTC

R topics documented:

  amzn_aapl_fb ........................................... 2
  proteus ............................................... 2

Index 6
amzn_aapl_fb  amzn_aapl_fb data set

Description
A data frame with the close prices for Amazon, Google and Facebook.

Usage
amzn_aapl_fb

Format
A data frame with 4 columns and 1798 rows.

Source
Yahoo Finance

proteus  proteus

Description
Seq2seq time-feature analysis based on variational model, with a wide range of distributions available for the latent variable.

Usage
proteus(
  data,
  target,
  future,
  past,
  ci = 0.8,
  deriv = 1,
  shift = 0,
  smoother = FALSE,
  t_embed = 30,
  activ = "linear",
  nodes = 32,
  distr = "normal",
  optim = "adam",
  loss_metric = "crps",
  epochs = 30,
  lr = 0.01,
proteus

```r
patience = 10,
verbose = TRUE,
seed = 42,
dev = "cpu",
dates = NULL,
dbreak = NULL,
days_off = NULL,
rolling_blocks = FALSE,
n_blocks = 4,
block_minset = 30,
batch_size = 30,
sequence_stride = FALSE
)
```

## Arguments

data A data frame with time features on columns and possibly a date column (not mandatory)

target Vector of strings. Names of the time features to be jointly analyzed

future Positive integer. The future dimension with number of time-steps to be predicted

past Positive integer. Length of past sequences

.ci Positive numeric. Confidence interval. Default: 0.8

deriv Positive integer or vector. Number of recursive differentiation operations for each time feature: for example, c(2, 1, 3) means the first feature will be differentiated two times, the second only one, the third three times. Default: 1 for each time feature.

shift Vector of positive integers. Allow for target variables to shift ahead of time. Zero means no shift. Length must be equal to the number of targets. Default: 0.

smoother Logical. Perform optimal smoothing using standard loess for each time feature. Default: FALSE

t Embed Positive integer. Number of embedding for the temporal dimension. Minimum value is equal to 2. Default: 30.


distr String. Distribution to be used by variational model. Implemented distributions are: "normal", "genbeta", "gev", "gpd", "genray", "cauchy", "exp", "logis", "chisq", "gumbel", "laplace", "lognorm". Default: "normal".

optim String. Optimization method. Implemented methods are: "adadelta", "adagrad", "rmsprop", "rprop", "sgd", "asgd", "adam".

loss_metric String. Loss function for the variational model. Two options: "elbo" or "crps". Default: "crps".

ePOCHS Positive integer. Default: 30.
lr

Positive numeric. Learning rate. Default: 0.01.

patience


verbose

Logical. Default: TRUE

seed

Random seed. Default: 42.

dev

String. Torch implementation of computational platform: "cpu" or "cuda" (gpu). Default: "cpu".

dates

Vector of strings. Vector with date strings for computing the prediction dates. Default: NULL (progressive numbers).

dbreak

String. Minimum time marker for x-axis plot, in liberal form: i.e., "3 months", "1 week", "20 days". Default: NULL.

days_off

String. Weekdays to exclude (i.e., c("saturday", "sunday")). Default: NULL.

rolling_blocks

Logical. Option for incremental or rolling window. Default: FALSE.

n_blocks


block_minset

Positive integer. Minimum number of sequence to create a block. Default: 30.

batch_size


sequence_stride

Logical. When FALSE, each sequence will be shifted of a single position in time; when TRUE, each sequence will be shifted for the full length of past + future (only distinct sequences allowed during reframing). Default: FALSE.

Value

This function returns a list including:

• prediction: a table with quantile predictions, mean, std, mode, skewness and kurtosis for each time feature
• plot: graph with history and prediction for each time feature
• learning_error: train and test error for the joint time features (rmse, mae, mdae, mpe, mape, smape, rrse, rae)
• feature_errors: train and test error for each time feature (rmse, mae, mdae, mpe, mape, smape, rrse, rae)
• pred_stats: for each predicted time feature, IQR to range, Kullback-Leibler Divergence (compared to previous point in time), upside probability (compared to previous point in time). Average for all the prediction statics and comparison between the terminal and the first point in the prediction sequence.
• time_log: computation time.

Author(s)

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

See Also

Useful links:

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

Examples

proteus(amzn_aapl_fb, c("AMZN", "GOOGL", "FB"), future = 30, past = 100)
proteus(amzn_aapl_fb, "AMZN", future = 30, past = 100, distr = "logis")
proteus(amzn_aapl_fb, "AMZN", future = 30, past = 100, distr = "cauchy")
proteus(amzn_aapl_fb, "AMZN", future = 30, past = 100, distr = "gev")
Index

* datasets
  amzn_aapl_fb, 2

amzn_aapl_fb, 2

proteus, 2
proteus-package (proteus), 2