

Package ‘argo’

November 16, 2015

Type Package

Title (ARGO) AutoRegression with GOogle search data - accurate estimation of influenza epidemics

Version 1.0

Date 2015-11-14

Author Shihao Yang

Maintainer Shihao Yang <shihaoyang@g.harvard.edu>

Description

Supplementary R package for PNAS article “Accurate estimation of influenza epidemics using Google search data via ARG0”, by Shihao Yang, Mauricio Santillana, and S. C. Kou.

License GPL-2

LazyData TRUE

Imports xts, glmnet, zoo, boot

Suggests testthat

NeedsCompilation no

R topics documented:

argo	2
bootstrap_relative_efficiency	3
heatmap_argo	3
load_data	4
logit	5
logit_inv	5
plot_argo	5
summary_argo	6
Index	7

argo

Construct ARGO object

Description

Wrapper for ARGO. The real work horse is glmnet package and/or linear model.

Usage

```
argo(data, exogen = xts::xts(NULL), N_lag = 1:52, N_training = 104,  
      alpha = 1, use_all_previous = FALSE)
```

Arguments

data	response variable as xts, last element can be NA
exogen	exogenous predictors, default is NULL
N_lag	vector of the AR model lags used, if NULL then no AR lags will be used
N_training	number of training points, if use_all_previous is true, this is the least number of training points required
alpha	penalty between lasso and ridge, alpha=1 represents lasso, alpha=0 represents ridge, alpha=NA represents no penalty
use_all_previous	boolean variable indicating whether to use "all available data" or "a sliding window" for training

Details

This function takes the time series and exogenous variables (optional) as input, and produces out-of-sample prediction for each time point.

Value

A list of following named objects

- pred An xts object with the same index as input, which contains historical nowcast estimation
- coef A matrix contains historical coefficient values of the predictors.
- parm Parameter values passed to argo function.

References

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. Proceedings of the National Academy of Sciences, doi: [10.1073/pnas.1515373112](https://doi.org/10.1073/pnas.1515373112).

```
bootstrap_relative_efficiency
      bootstrap relative efficiency confidence interval
```

Description

This function is used to reproduce the ARGO bootstrap confidence interval of PNAS paper.

Usage

```
bootstrap_relative_efficiency(pred_data, model_good, model_bench, l = 50,
  N = 10000, truth = "CDC.data", sim = "geom", conf = 0.95,
  type = c("mse", "mape", "mae"))
```

Arguments

pred_data	A matrix that contains the truth vector and the predictions. It can be data.frame or xts object
model_good	The model to evaluate, must be in the column names of pred_data
model_bench	The model to compare to, must be in the column names of pred_data
l	stationary bootstrap mean block length
N	number of bootstrap samples
truth	the column name of the truth
sim	simulation method, pass to boot::tsboot
conf	confidence level
type	Must be one of "mse" (mean square error), "mape" (mean absolute percentage error), or "mae" (mean absolute error)

Value

A vector of point estimate and corresponding bootstrap confidence interval

```
heatmap_argo      Heatmap plot of ARGO coefficients applied on CDC's ILI data
```

Description

This function is only used to reproduce the ARGO plot of PNAS paper. If you use this plotting routine for other dataset, an error is highly likely to occur.

Usage

```
heatmap_argo(argo_coef, lim)
```

Arguments

argo_coef	The coefficient matrix
lim	the limit to truncate for large coefficients for better presentation

Value

a graph on the default plot window

load_data

Parsing of raw data

Description

Data related to the PNAS paper. Accessed on Nov 14, 2015.

Usage

```
load_data()
```

Details

Parse and load CDC's ILI data, Google Flu Trend data, Google Correlate data trained with ILI as of 2010, Google Correlate data trained with ILI as of 2009, Google Trend data with search terms identified from Google Correlate (2010 version).

Each week ends on the Saturday indicated in the xts object

Google Correlate data is standardized by Google, and we rescale it to 0 – 100 during parsing. Google Trends data is in the scale of 0 – 100.

Value

A list of following named xts objects

- GC10 Google Correlate trained with ILI available as of 2010. Available online at <https://www.google.com/trends/correlate/search?e=id:20xKcnNqHrk&t=weekly>
- GC09 Google Correlate trained with ILI available as of 2009. Not directly available online, you have to manually input ILI time series at <https://www.google.com/trends/correlate>
- GT Google Trends data for search queries identified using Google Correlate. Not directly available online, you have to manually input query terms at <https://www.google.com/trends>
- CDC CDC's ILI dataset. Available online at <http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>
- GFT Google Flu Trend (historical predictions). Available online at <https://www.google.org/flutrends>

References

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. Proceedings of the National Academy of Sciences, doi: [10.1073/pnas.1515373112](https://doi.org/10.1073/pnas.1515373112).

Examples

```
system.file("extdata", "correlate-Influenza_like_Illness_h1n1_CDC_.csv", package = "argo")
system.file("extdata", "correlate-Influenza_like_Illness_CDC_.csv", package = "argo")
system.file("extdata", "GFT.csv", package = "argo")
system.file("extdata", "ILINet.csv", package = "argo")
load_data()
```

logit	<i>logit function</i>
-------	-----------------------

Description

logit function

Usage

logit(x)

logit_inv	<i>inverse logit function</i>
-----------	-------------------------------

Description

inverse logit function

Usage

logit_inv(x)

plot_argo	<i>Time series plot of ARGO applied on CDC's ILI data</i>
-----------	---

Description

This function is only used to reproduce the ARGO plot of PNAS paper. If you use this plotting routine for other dataset, an error is highly likely to occur.

Usage

plot_argo(GFT_xts, GC_GT_cut_date, model_names, legend_names, zoom_periods)

Value

a graph on the default plot window

`summary_argo`*performance summary of ARGO applied on CDC's ILI data*

Description

This function is only used to reproduce the ARGO table of PNAS paper. If you use this summary function for other dataset, an error is highly likely to occur.

Usage

```
summary_argo(GFT_xts, model_names, legend_names, periods,  
             whole_period = "2009-03/2015-10")
```

Value

A list of summary tables for the inputted periods, including RMSE, MAE, MAPE, corr

Examples

```
summary_argo
```

Index

argo, [2](#)

bootstrap_relative_efficiency, [3](#)

heatmap_argo, [3](#)

load_data, [4](#)

logit, [5](#)

logit_inv, [5](#)

plot_argo, [5](#)

summary_argo, [6](#)