Package ‘theft’

March 11, 2024

Type Package

Title Tools for Handling Extraction of Features from Time Series

Version 0.6.1

Date 2024-03-11

Maintainer Trent Henderson <then6675@uni.sydney.edu.au>

Description Consolidates and calculates different sets of time-series features from multiple 'R' and 'Python' packages including 'Rcatch22' Henderson, T. (2021) <doi:10.5281/zenodo.5546815>, 'feasts' O'Hara-Wild, M., Hyndman, R., and Wang, E. (2021) <https://CRAN.R-project.org/package=feasts>, 'tsfeatures' Hyndman, R., Kang, Y., Montero-Manso, P., Talagala, T., Wang, E., Yang, Y., and O'Hara-Wild, M. (2020) <https://CRAN.R-project.org/package=tsfeatures>, 'tsfresh' Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr A.W. (2018) <doi:10.1016/j.neucom.2018.03.067>, 'TSFEL' Barandas, M., et al. (2020) <doi:10.1016/j.softx.2020.100456>, and 'Kats' Facebook Infrastructure Data Science (2021) <https://facebookresearch.github.io/Kats/>.

BugReports https://github.com/hendersontrent/theft/issues

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports utils, rlang, dplyr, tidyr, purrr, tibble, tsibble, fabletools, tsfeatures, feasts, Rcatch22, reticulate, R.matlab

Suggests theftdlc, lifecycle, cachem, bslib, knitr, markdown, rmarkdown, pkgdown, testthat

RoxygenNote 7.2.2

VignetteBuilder knitr

URL https://hendersontrent.github.io/theft/

NeedsCompilation no

Author Trent Henderson [cre, aut],
Annie Bryant [ctb]

Repository CRAN

Date/Publication 2024-03-11 11:20:02 UTC

R topics documented:

`calculate_features` *Compute features on an input time series dataset*

Description

Compute features on an input time series dataset

Usage

```
calculate_features(
  data,
  id_var = "id",
  time_var = "timepoint",
  values_var = "values",
  group_var = NULL,
 feature_set = c("catch22", "feasts", "tsfeatures", "Kats", "tsfresh", "TSFEL"),
  catch24 = FALSE,
  tsfresh_cleanup = FALSE,
  features = NULL,
  seed = 123
)
```

Arguments

`data`	`data.frame` with at least 4 columns: id variable, group variable, time variable, value variable

id_var	character specifying the ID variable to identify each time series. Defaults to "id"
time_var	character specifying the time index variable. Defaults to "timepoint"
values_var	character specifying the values variable. Defaults to "values"
group_var	character specifying the grouping variable that each unique series sits under (if one exists). Defaults to NULL
feature_set	character or vector of character denoting the set of time-series features to calculate. Defaults to "catch22"
catch24	Boolean specifying whether to compute catch24 in addition to catch22 if catch22 is one of the feature sets selected. Defaults to FALSE
tsfresh_cleanup	Boolean specifying whether to use the in-built tsfresh relevant feature filter or not. Defaults to FALSE
features	named list containing a set of user-supplied functions to calculate on data. Each function should take a single argument which is the time series. Defaults to NULL for no manually-specified features. Each list entry must have a name as calculate_features looks for these to name the features. If you don't want to use the existing feature sets and only compute those passed to features, set feature_set = NULL
seed	integer denoting a fixed number for R's random number generator to ensure reproducibility. Defaults to 123

Value

object of class feature_calculations that contains the summary statistics for each feature

Author(s)

Trent Henderson

Examples

```
featMat <- calculate_features(data = simData,
  id_var = "id",
  time_var = "timepoint",
  values_var = "values",
  group_var = "process",
  feature_set = "catch22",
  seed = 123)
```

check_vector_quality *Check for presence of NAs and non-numerics in a vector*

Description

Check for presence of NAs and non-numerics in a vector

Usage

```
check_vector_quality(x)
```

Arguments

x	input vector

Value

Boolean of whether the data is good to extract features on or not

Author(s)

Trent Henderson

feature_list *All features available in theft in tidy format*

Description

The variables include:

Usage

```
feature_list
```

Format

A tidy data frame with 2 variables:

feature_set Name of the set the feature is from

feature Name of the feature

`init_theft` *Communicate to R the Python virtual environment containing the relevant libraries for calculating features*

Description

Communicate to R the Python virtual environment containing the relevant libraries for calculating features

Usage

```
init_theft(venv)
```

Arguments

`venv` `character` specifying the name of the to the Python virtual environment where `"tsfresh"`, `"TSFEL"`, and/or `"Kats"` are installed

Value

no return value; called for side effects

Author(s)

Trent Henderson

Examples

```
## Not run:
install_python_pkgs("theft-test")
init_theft("theft-test")

## End(Not run)
```

`install_python_pkgs` *Download and install all the relevant Python packages into a target location*

Description

Download and install all the relevant Python packages into a target location

Usage

```
install_python_pkgs(venv, standard_kats = TRUE)
```

Arguments

`venv` `character` specifying the name of the new virtual environment to create

`standard_kats` `Boolean` denoting whether to try a standard installation of `Kats` from PyPI using `reticulate::virtualenv_install` or to install a safer version with less dependencies. Defaults to `TRUE`

Value

no return value; called for side effects

Author(s)

Trent Henderson

Examples

```
## Not run:
install_python_pkgs("theft-test")

## End(Not run)
```

`process_hctsa_file` *Load in hctsa formatted MATLAB files of time series data into a tidy format ready for feature extraction*

Description

Load in hctsa formatted MATLAB files of time series data into a tidy format ready for feature extraction

Usage

```
process_hctsa_file(data)
```

Arguments

`data` `string` specifying the filepath to the `MATLAB` file to parse

Value

an object of class `data.frame` in tidy format

Author(s)

Trent Henderson

`simData` *Sample of randomly-generated time series to produce function tests and vignettes*

Description

The variables include:

Usage

```
simData
```

Format

A tidy data frame with 4 variables:

id Unique identifier for the time series

timepoint Time index

values Value

process Group label for the type of time series

`theft` *Tools for Handling Extraction of Features from Time-series*

Description

Tools for Handling Extraction of Features from Time-series

Index