Package 'UpDown'

July 20, 2023

Type Package

Title Detecting Group Disturbances from Longitudinal Observations

Version 1.2.1

Description Provides an algorithm to detect and characterize disturbances (start, end dates, intensity) that can occur at different hierarchical levels by studying the dynamics of longitudinal observations at the unit level and group level based on Nadaraya-Watson's smoothing curves, but also a shiny app which allows to visualize the observations and the detected disturbances. Finally the package provides a dataframe mimicking a pig farming system subsected to disturbances simulated according to Le et al.(2022) <doi:10.1016/j.animal.2022.100496>.

License GPL (>= 3)

Encoding UTF-8

LazyDataCompression xz

Imports stats, mixtools, mclust, dplyr, ggplot2, reshape2, shiny

Depends R (>= 4.0.0)

LazyData true

Author Tom Rohmer [aut, cre],
Vincent Le [aut],
Ingrid David [aut]

Maintainer Tom Rohmer <tom.rohmer@inrae.fr>

NeedsCompilation no

Repository CRAN

Date/Publication 2023-07-20 18:50:13 UTC

R topics documented:

PigDisturbance *Non-observable information about the disturbances occuring in the* PigFarming *dataset*

Description

Information about the disturbances relative to the dataset PigFarming: start, and intensity. An intensity of 0 stands for no disturbance for the considered hierarchical level. Not observable in practice.

Usage

```
data("PigDisturbance")
```

Format

A dataframe with 6000 records on the following 12 variables:

id the identifier of the animal

batch the numero of the batch of the animal

pen the numero of the batch of the animal

int_batch the intensity of the disturbance at the batch level

start_batch the starting time of the disturbance at the batch level

end_batch the endding time of the disturbance at the batch level

int_pen the intensity of the disturbance at the pen level

start_pen the starting time of the disturbance at the pen level

end_pen the endding time of the disturbance at the pen level

int_ind the intensity of the disturbance at the individual level

start_ind the starting time of the disturbance at the individual level

end_ind the endding time of the disturbance at the individual level

References

Le, Vincent, Tom Rohmer, and Ingrid David. 2022. "Impact of Environmental Disturbances on Estimated Genetic Parameters and Breeding Values for Growth Traits in Pigs." Animal 16 (4): 100496. https://doi.org/10.1016/j.animal.2022.100496

Examples

```
theo_dat=get(data(PigDisturbance))
str(theo_dat)
theo_dat[c(1,6,5405),]
```

PigFarming *Simulated longitudinal phenotypes that mimics a pig-farming dataset subsected to disturbances*

Description

Example of a dataset on which UpDown can be applied. It consists in simulated hierarchical data mimicking a pig farming-system dataset subsected to disturbances. The animals (id) were raised in 40 batches and in 15 pens within each batch leading to 15 animals per pen. Hence three hierarchical level are considered: id, pen and batch levels. Data were simulated following Le et al. 2022. <doi.org/10.1016/j.animal.2022.100496>

Usage

```
data("PigFarming")
```

Format

A data frame with 578847 individual observations on the following 6 variables:

`id` the identifier of the animal

`batch` the numero of the batch

`pen` the numero of the pen

`age` the age (in day) of the animal

`time` the observation times

`weight` the weight (in kg) of the animal

References

Le Vincent, Tom Rohmer, and Ingrid David. 2022. "Impact of Environmental Disturbances on Estimated Genetic Parameters and Breeding Values for Growth Traits in Pigs." Animal 16 (4): 100496. https://doi.org/10.1016/j.animal.2022.100496

Examples

```
data=get(data(PigFarming))
str(data)
plot(subset(data,id==6)$weight)
```

UpDown *Unsupervised hierarchical classification for elastic or plastic response to a disturbance using mixture of Gaussian distributions*

Description

Detection and characterisation of disturbances from longitudinal data, organized in hierarchical groups

Usage

```
UpDown(data,levels,obs, vtime, h.int=NULL,
mixplot=FALSE, correction=NULL,
kappa=NULL,thr_va=0.5,
options=list())
```

Arguments

`data`	a dataframe containing at minima observations, a time variable and hierarchical levels (one column per level). The dataframe can also contain other variables. One row per unit and observed time is needed.
`levels`	a vector of character strings specifing the column names corresponding to the considered hierarchical levels appearing in the dataframe `data`, ordered from the highest hierarchical level to the lowest hierarchical level (i.e., the unit level)
`obs`	character string specifing the column names of the considered numeric observations appearing in the data frame `data`.
`vtime`	character string specifing the column names of the considered time variable appearing in the data frame `data`.
`h.int`	a real parameter specifying the smoothing bandwidth in Nadaraya-Watson's smoothing curves *see ?ksmooth*. The default value is $\sqrt{n}$, where n is the largest length of observations per unit.
`mixplot`	logical. If TRUE, the mixture curves for each hierarchical levels are plotted. (default value FALSE)
`correction`	an optional character string specifing the column name of a considered time-dependent discrete variable appearing in the dataframe `data` for which a correction of the observation `obs` is needed. The correction consists in subtracting the median relative to the modalities of the variable.
`kappa`	an integer in [0,1) used to eliminate a redundancy disturbance between two distinct levels, based on the estimated starting and end points. It evaluates the overlapping between two considered disturbances. The disturbance is removed in the lowest of the two hierarchical levels. When kappa is not specified (null default value) no redounding disturbances are eliminated. When kappa is equal to 1, only the disturbances with the same rounded starting and ending points are removed. An excessively small values for kappa can lead to wrongly remove disturbances. The suggested value for kappa is 0.75.

`thr_va` an integer in [0,1) used to validate in the down step the group disturbances. `thr_va` control the percentage of starting-times of the underlying elements in a close interval to validate the disturbance. A value of 0 indicate that all the disturbances are validated that is to say no validation is done. The default value is 0.5, i.e., at least 50% of the elements constituting the group should have a disturbance in a close time to validate the group disturbance.

`options` A list of options. See the documentation *see ?normalmixEM* for possible options of the mixture model

Details

Note that unique identifiers are mandatory for the hierarchical levels. Moreover note that UpDown considers that all perturbations have a negative effect on the longitudinal observations. For a positive effect, consider the opposite sign of the observations before using UpDown. Units with less than 20 observations are removed. That can be modified using the option `minobs` in `options`.

Value

A list containing the following components:

`data` the initial dataframe with the supplementary columns (if `correction` is not null); *medobs*, the median of observations grouped into the modalities of the qualitative variable used in `correction`; and *cobs* the corrected observations. Units with less than `minobs` observations are removed.

`levels` the specified hierarchical levels.

`med_lev` a list of matrices. The last matrix contains the medians of the unit observations per observed time. The previous one contains the medians of these medians per observed time and so on up to the highest hierarchical level. If there are less than 50% observations per considered time, the median will not be evaluated (NA).

`mixmdl_lev` a list of outputs of the mixture models of the hierarchical levels. The first output concerns the lowest hierarchical levels (i.e., unit) and the last output concerns the highest hierarchical levels.

`names_lev` a list of names of each elements of the hierarchical levels. The first matrix concerns the lowest hierarchical levels (i.e., unit) and the last matrix concerns the highest hierarchical levels.

`sc.x_lev` a list of matrices giving the time points considered by the smoothing, per identifier and for each hierarchical level. The first matrix concerns the lowest hierarchical levels (i.e., unit) and the last matrix concerns the highest hierarchical levels.

`sc.y_lev` a list of matrices of fitted values corresponding to `sc.x_lev`. The first matrix concerns the lowest hierarchical levels (i.e., unit) and the last matrix concerns the highest hierarchical levels.

`sc.dx_lev` a list of matrices giving the time points considered by the derivative smoothing curve, per identifier and for each hierarchical level. The first matrix concerns the lowest hierarchical levels (i.e., unit) and the last matrix concerns the highest hierarchical levels.

sc.dy_lev	a list of matrices of fitted values corresponding to sc.dx_lev. The first matrix concerns the lowest hierarchical levels (i.e., unit) and the last matrix concerns the highest hierarchical levels.
Up	a dataframe that describes for each ids, the type of detected disturbance at the end of the Up-step. '0' stands for no disturbance.
Down	a list of matrices that gives the detected disturbances for each hierarchical level and their characteristics.

Author(s)

Tom Rohmer, Vincent Le, Ingrid David

References

Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. mixtools: An R package for analyzing finite mixture models. Journal of Statistical Software, 32(6):1-29, 2009.

Le, V. 2022. "Nouvelle mesure de la robustesse des animaux d'élevage par utilisation des données de phénotypage haut-débit." Thesis, INPT Toulouse. https://hal.inrae.fr/tel-03967884.

Nadaraya, E. A. On estimating regression. Theory of Probability & Its Applications, 9(1):141–142, 1964

Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models, The R Journal, 8/1, pp. 289-317.

Watson, G. S. Smooth regression analysis. Sankhya: The Indian Journal of Statistics, Series A, pages 359–372, 1964

See Also

normalmixEM(), mclust(), ksmooth()

Examples

```
# load data
data.ex=get(data(PigFarming))

# optional arguments
options<-list(maxit=100)

# considered hierarchical levels
levels=c("batch","pen","id")

UpDown.out<- UpDown(data.ex, levels=levels, vtime="time", obs="weight",
kappa=0.75, thr_va=0.5, h.int=10, mixplot=FALSE, correction="age", options=options)

UpDown.out$Down$batch
```

UpDownApp *Shiny app to vizualise the longitudinal intra-group observations and the detected disturbances*

Description

The function starts a shiny app which visualizes the data organized by the hierarchical levels, and the estimated start and end point of the detected disturbances

Usage

```
UpDownApp(updown.out,obs=NULL,width=1000,height=1000)
```

Arguments

updown.out	Global output of the UpDown function
obs	(optional) vector of character string specifing the names of the considered longitudinal observations appearing in the dataframe data.
width,height	(optional) interger specifing the width and the height of the plot in the Rshiny.

Value

No return value, run the application

See Also

shiny(), ggplot2(), UpDown()

Examples

```
# load data
data=get(data(PigFarming))

levels=c("batch","pen","id")
updown.out<- UpDown(data,levels=levels, vtime="time", obs="weight",
kappa=0.75, thr_va=0.5, correction="age")


if(interactive()){UpDownApp(updown.out)}
```

Index