Package 'PML'

September 17, 2019

Type Package

Title Penalized Multi-Band Learning for Circadian Rhythm Analysis Using Actigraphy

Version 1.1

Date 2019-09-17

Maintainer Xinyue Li <xinyue.li@yale.edu>

Description Penalized Multi-Band Learning algorithm can be effectively implemented for circadian rhythm analysis and daily activity pattern characterization using actigraphy (continuously measured objective physical activity data). Functions for interactive visualization of actigraph data are also included. Method reference: Li, X., Kane, M., Zhang, Y., Sun, W., Song, Y., Dong, S., Lin, Q., Zhu, Q., Jiang, F., Zhao, H. (2019) A Novel Penalized Multi-band Learning Approach Characterizes the Consolidation of Sleep-Wake Circadian Rhythms During Early Childhood Development.

Depends R (>= 3.4.0)

License GPL (>= 2)

Imports trelliscopejs,tidyr,rbokeh,dplyr,tibble

Encoding UTF-8

Suggests knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 6.1.1

URL https://github.com/xinyue-L/PML

BugReports https://github.com/xinyue-L/PML/issues

NeedsCompilation no

Author Xinyue Li [aut, cre], Michael Kane [aut]

Repository CRAN

Date/Publication 2019-09-17 08:00:02 UTC

R topics documented:

`PML-package` *Penalized Multi-Band Learning for Circadian Rhythm Analysis Using Actigraphy*

Description

Penalized Multi-Band Learning algorithm can be effectively implemented for circadian rhythm analysis and daily activity pattern characterization using actigraphy (continuously measured objective physical activity data). Functions for interactive visualization of actigraph data are also included. Method reference: Li, X., Kane, M., Zhang, Y., Sun, W., Song, Y., Dong, S., Lin, Q., Zhu, Q., Jiang, F., Zhao, H. (2019) A Novel Penalized Multi-band Learning Approach Characterizes the Consolidation of Sleep-Wake Circadian Rhythms During Early Childhood Development.

Details

The DESCRIPTION file:

Package:	PML
Type:	Package
Title:	Penalized Multi-Band Learning for Circadian Rhythm Analysis Using Actigraphy
Version:	1.1
Date:	2019-09-17
Authors@R:	c(person(given = "Xinyue", family = "Li", email = "xinyue.li@yale.edu", role = c("aut", "cre")), person(
Maintainer:	Xinyue Li <xinyue.li@yale.edu>
Description:	Penalized Multi-Band Learning algorithm can be effectively implemented for circadian rhythm analysis
Depends:	R (>= 3.4.0)
License:	GPL (>= 2)
Imports:	trelliscopejs,tidyr,rbokeh,dplyr,tibble
Encoding:	UTF-8
Suggests:	knitr, rmarkdown
VignetteBuilder:	knitr
RoxygenNote:	6.1.1
URL:	https://github.com/xinyue-L/PML

BugReports:	https://github.com/xinyue-L/PML/issues
NeedsCompilation:	no
Packaged:	2019-09-17 06:23:33 UTC; xinyueli
Author:	Xinyue Li [aut, cre], Michael Kane [aut]

Index of help topics:

Penalized Multi-Band Learning algorithm can be effectively implemented for circadian rhythm analysis and daily activity pattern characterization using actigraphy (continuously measured objective physical activity data). Functions for interactive visualization of actigraph data are also included.

Author(s)

Xinyue Li [aut, cre], Michael Kane [aut]

Maintainer: Xinyue Li <xinyue.li@yale.edu>

References

Li, X., Kane, M., Zhang, Y., Sun, W., Song, Y., Dong, S., Lin, Q., Zhu, Q., Jiang, F., Zhao, H. (2019) A Novel Penalized Multi-band Learning Approach Characterizes the Consolidation of Sleep-Wake Circadian Rhythms During Early Childhood Development.

Fisher, R. A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society of London. Series A, 125(796), 54-59.

Examples

```
library(PML)
##reformat data for further analysis
data(lis3)
```

```
pa3 <- form(lis3)

##apply Penalized Multi-band Learning
data(pa3)
re <- bandSelect(df=pa3,Nlength=1440*3,Nlambda=100,alpha=1,Ntop=3,cross=FALSE,Ncross=NULL,plot=TRUE)

##use trelliscope to visualize data:
##return a dataset with trelliscope panels for individual mean activity plots
data(var3)
tre.ind <- tre(lis3,varlis=var3)
tre.ind$activity_ind <- tre.ind$activity_all <- NULL
```

---

bandSelect *Penalized multi-band learning function*

Description

In a group of individuals with physcial activity data, this function utilize Fast Fourier Transform (FFT) and L-1/L-2 penalties to select significant harmonics/periodicities and describe the main activity pattern (circadian rhythm) among the population.

Usage

```
bandSelect(df, Nlength, Nlambda = 100, alpha = 1, Ntop = 5,
  cross = FALSE, Ncross = NULL, plot = TRUE)
```

Arguments

df	the tbl_df data frame containing at least two variables: subject ID and activity. The function form can help prepare the data frame.
Nlength	the length of observations necessary for each individual, note that it should be consistent among all
Nlambda	λ's take values from 0 to 2max(‖X_k‖^2), as 0 gives no penalty and the latter suppresses all θ's to 0. Therefore, we divide 2max(‖X_k‖^2) into Nlambda (default to be 100) λ's to pick frequencies/harmonics/periodicities.
alpha	the tuning parameter controlling the balance between L-1 and L-2 penalty. The default is 1, using complete Lasso/ L-1 penalty.
Ntop	the number of frequencies/harmonics/periodicities picked for the population. The default is 5.
cross	whether to perform cross-validation. The default is FALSE.
Ncross	the number of groups of data for cross-validation. If cross=TRUE, the data shall be divided into Ncross groups.
plot	whether to plot: MSE against the number of nonzero θ's, and only the points at which the number of nonzero θ's changes (as λ changes) are be plotted. The default is TRUE.

Value

if no cross-validation is conducted, return a list; if cross-validation, return a list of lists, with the last list consisting of all FFT results and cross-validation groups (showing the subject IDs leave-out /NOT used each time).

`topfreq`	vector of length `Ntop`: top frequencies selected.
`mse`	vector of length `Nlambda`: mean squared error for each lambda (penalty). If no cross-validation, mse is calculated based on all available data; if cross-validation, mse is calculated based on the rest observations.
`nonzero`	vector of length `Nlambda`: the number of nonzero θ's (frequencies) for each lambda (penalty).
`deltazero`	vector of length `Nlambda`: the change in the number of nonzero θ's (frequencies) for each lambda (penalty).
`lambda`	vector of length `Nlambda`: the value of lambda.
`theta`	`Nfreq` by `Nlambda` matrix: estimated θ's (frequencies) at each lambda (penalty). Nfreq is the total number of frequencies given by FFT.
`xscore`	`Nind` by `Nfreq` matrix: the original FFT scores for each individual. Nind is the number of individuals in the population, and Nfreq is the total number of frequencies given by FFT.
`xprop`	`Nind` by `Nfreq` matrix: the original FFT results expressed as the proportion of variances explained by each frequency for each individual. Nind is the number of individuals in the population, and Nfreq is the total number of frequencies given by FFT.
`freq`	vector of length `Nfreq`: list of frequencies in FFT results.

References

Li, X. , Kane, M. , Zhang, Y. , Sun, W. , Song, Y. , Dong, S. , Lin, Q. , Zhu, Q. , Jiang, F. & Zhao, H. (2019). Penalized Selection of Periodicities Characterizes the Consolidation of Sleep-Wake Circadian Rhythms During Early Childhood Development. Submitted.

See Also

`form`

Examples

```
data(pa3)
re <- bandSelect(df=pa3,Nlength=1440*3,Nlambda=100,alpha=1,Ntop=5,
                 cross=FALSE,Ncross=NULL,plot=TRUE)
```

`form` *Function to generate activity data frame for penalized multi-band learning*

Description

This function generates the data frame necessary for further penalized multi-band learning.

Usage

```
form(lis, maxday = 14, id = NULL)
```

Arguments

`lis`	the list of activity data, with each element corresponding to the observation by one individual and the name of each element coresponding to the individual id. Specifically, each element is a `nob` by `nday` matrix, where each column is an observation by day.
`maxday`	the maximal number of days per individual in the observation, used to check the data format. The default is 14.
`id`	a vector of id names corresponding to the `lis` activity data.

Value

The activity data frame with 3 columns: ID, IDday, and activity.

See Also

`bandSelect`

Examples

```
data(lis3)
pa3 <- form(lis3)
```

`gharmonic` *harmonic analysis test: g-value calculation*

Description

This function calculates the g-value for the harmonic analysis test developed by R.A. Fisher (1929). Harmonic analysis refers to Fast Fourier Transform (FFT) results. Specifically, g is the proportion (squared modulus of one frequency divided by the sum of all squared moduli). In order for g to be statistically significant in the harmonic analysis test, it needs to be at least g-value at significance level α. Please note that for the rth largest frequency, if any of the previous (r-1) frequencies is not significant, then the rth largest frequency is also non-significant.

Usage

```
gharmonic(n, r, p, tol = 10^-7, init = NULL)
```

Arguments

n	the total number of frequencies in FFT results.
r	the modulus of the tested frequency is ranked as the rth largest among all frequencies.
p	the FFT result of the tested frequency expressed as the squared modulus divided by the sum of the squared moduli by all frequencies (proportion: m_r^2/(m_1^2+...+m_n^2)).
tol	the tolerance level during calculation. The default is 10^-7.
init	the crude estimate for g-value if known. It is not called to calculate usual g-values.

Value

The g-value calculated by the harmonic test.

References

Fisher, R. A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society of London. Series A, 125(796), 54-59.

See Also

pharmonic

Examples

```
gharmonic(n=100,r=1,p=0.05)
```

---

lis3 *An example of individual activity data*

Description

'lis3' is a data list consisting of three matrices, each giving activity data for one individual. Each column of the matrix is one-day observation, and here the physical activity (PA) is measured every one minute, so the matrix is 1440 by 'nday'. 'lis3' is also a named list, the name of which is the individual ID.

Usage

```
data(lis3)
```

Format

A data list of activity data for 3 individuals; each inidivdual data set is a 1440 by nday matrix.

`pa3` *An example of reformated individual activity data*

Description

'pa3' is in tbl_df format. It has 13 observations for 3 individuals, and the variables are "ID", "ID_Nday" (the ith day observation for an individual), and activity. The activity variable is an embedded list with each element consisting of a vector of one-day observation.

Usage

```
data(pa3)
```

Format

tbl_df

`pharmonic` *Harmonic analysis test: p-value calculation*

Description

This function calculates the p-value for the harmonic analysis test developed by R.A. Fisher (1929). Harmonic analysis specifically refers to Fast Fourier Transform (FFT) results.

Usage

```
pharmonic(n, r, g)
```

Arguments

`n`	the total number of frequencies in FFT results
`r`	the modulus of the tested frequency is ranked as the rth largest among all frequencies
`g`	the FFT result of the tested frequency expressed as the squared modulus divided by the sum of the squared moduli by all frequencies (proportion: m_r^2/(m_1^2+...+m_n^2)).

Value

The p-value calculated by the harmonic test.

References

Fisher, R. A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society of London. Series A, 125(796), 54-59.

See Also

gharmonic

Examples

```
pharmonic(n=100,r=2,g=0.1)
```

`test.harmonic` *Harmonic analysis test for Fast Fourier Transform*

Description

This function conducts harmonic test sequentially based on observations or Fast Fourier Transform (FFT) results.

Usage

```
test.harmonic(ob, p, fft = FALSE, maxfreq = 10)
```

Arguments

`ob`	Either the original observation or FFT results. See parameter `fft`.
`p`	The p-value to be considered statistically significant.
`fft`	If TRUE, `ob` is FFT results, with the first column frequencies and the second column signals in standardized proportions; if FALSE, `ob` is a vector of the original observation. The default is FALSE.
`maxfreq`	To conduct test on at most `maxfreq` frequencies. The default is 10.

Value

A list of two elements:

`sig`	The significant frequencies plus the first insignificant frequency.
`fft`	The FFT results expressed in standardized proportions.

References

Fisher, R. A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society of London. Series A, 125(796), 54-59.

See Also

pharmonic

Examples

```
data(pa3)

#### test on individuals
ob <- do.call("c",pa3$activity[1:4])
re <- test.harmonic(ob,p=0.05/(length(ob)-1)/2)
re$sig;head(re$fft) ## no harmonic is significant
ob2 <- do.call("c",pa3$activity[11:13])
re2 <- test.harmonic(ob2,p=0.05/(length(ob2)-1)/2)
re2$sig;head(re2$fft) ## 3 significant harmonics

#### test on the population average
re0 <- bandSelect(df=pa3,Nlength=1440*3,Nlambda=100,alpha=1,Ntop=3,
                  cross=FALSE,Ncross=NULL,plot=TRUE)
freq <- data.frame(Frequency=re0$freq,Proportion=colMeans(re0$xprop))
re3 <- test.harmonic(freq,p=0.05/nrow(freq),fft=TRUE)
print(re3$sig,digits=3,row.names=FALSE)
```

---

tre *Trelliscope Visualization for Accelerometer Data*

Description

This function generates the data frame necessary for trelliscope visualization.

Usage

```
tre(lis, id = NULL, varlis = NULL, smband = 1/12, maxday = 14,
  plot.ind = TRUE, plot.ori = TRUE, plot.sm = TRUE,
  plot.tre = FALSE, plot.tre.path = NULL)
```

Arguments

lis	the list of activity data, with each element corresponding to the observation by one individual and the name of each element coresponding to the individual id. Specifically, each element is a nob by nday matrix, where each column is an observation by day.
id	a vector of id names corresponding to the lis activity data.
varlis	optional data frame to be merged to activity data, and the covariates are of interest for plotting to see activity differences. The first variables needs to be "ID".
smband	smoothing parameter for plotting smoothed activity data. the default is 1/12 (see function lowess).

maxday	maxday the maximal number of days per individual in the observation, used to check the data format. The default is 14.
plot.ind	whether to plot individual mean activity plots. If not, plot day activity plots. The default is TRUE.
plot.ori	whether to plot the original activity curves (tend to have large variations). The default is TRUE.
plot.sm	whether to plot lowess of the activity curves. The default is TRUE.
plot.tre	whether to generate trelliscope plots. If so, no data will be returned; if not, a data frame will be returned containing all information including trelliscope panels. To generate trelliscope based on the data, one needs to set all activity list columns to NULL. The default is FALSE.
plot.tre.path	If plot.tre is TRUE, then plot.tre.path specifies the path to generate trelliscope files. The default is current working directory.

Value

The data frame including activity, filtering stats, optional covariates, and trelliscope panels. (No data frame will be returned if plot.tre is TRUE.)

See Also

form

Examples

```
data(lis3)
data(var3)

#### individual mean activity plot: return a dataset with trelliscope panels
tre.ind <- tre(lis3,varlis=var3)
tre.ind$activity_ind <- tre.ind$activity_all <- NULL
```

---

var3 *Demographic information for individuals in datasets lis3 and pa3*

Description

'var3' is a data frame consisting of ID, gender, and age information for the 3 individuals in datasets 'lis3' and 'pa3'. 'var3' is used as an input to generate trelliscope panels for visualization so that one can filter based on age and gender.

Usage

```
data(var3)
```

Format

a data frame consisting of three variables: ID, gender and age

Index