Package ‘Bagidis’

June 26, 2015

Type Package

Title BAses GIving DIStances

Version 1.0

Date 2015-06-24

Author Catherine Timmermans

Maintainer Catherine Timmermans <catherine.timmermans@uclouvain.be>

Description This is the companion package of a PhD thesis entitled ``Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy'' by Timmermans (2012). See references for details and related publications. The core of the BAGIDIS methodology is a functional wavelet based semi-distance that has been introduced by Timmermans and von Sachs (2010, 2015) and Timmermans, Delsol and von Sachs (2013). This semi-distance allows for comparing curves with sharp local patterns that might not be well aligned from one curve to another. It is data-driven and highly adaptive to the curves being studied. Its main originality is its ability to consider simultaneously horizontal and vertical variations of patterns, which proofs highly useful when used together with clustering algorithms or visualization method. BAGIDIS is an acronym for BAsis GIving DIStances. The extension of BAGIDIS to image data relies on the same principles and has been described in Timmermans and Fryzlewicz (2012), Fryzlewicz and Timmermans (2015).

License GPL

Depends abind

NeedsCompilation no

Repository CRAN

Date/Publication 2015-06-26 00:36:18

R topics documented:

`BAGIDIS-package` *BAGIDIS (BAses Giving DIStances). A New Set of Tools for Investigating Functional Data With Sharp Local Patterns*

Description

This is the companion package of a PhD thesis entitled "Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy" by *Timmermans (2012)*. See references below for details and related publications.

The core of the BAGIDIS methodology is a functional wavelet based semi-distance that has been introduced by *Timmermans and von Sachs (2010, 2015) and Timmermans, Delsol and von Sachs (2013)*. This semi-distance allows for comparing curves with sharp local patterns that might not be well aligned from one curve to another. It is data-driven and highly adaptive to the curves being studied. Its main originality is its ability to consider simultaneously horizontal and vertical variations of patterns, which proofs highly useful when used together with clustering algorithms or visualization method. BAGIDIS is an acronym for *BAsis GIving DIStances*. The extension of BAGIDIS to image data relies on the same principles and has been described in *Timmermans and Fryzlewicz (2012), Fryzlewicz and Timmermans (2015)*.

Details

Package:	Bagidis
Type:	Package
Version:	2.0
Date:	2012-05-03
License:	GPL-3
LazyLoad:	yes

The BAGIDIS methodology aims at answering the need for a method able to detect the closeness of curves or images whose significant sharp features might not be well aligned. It has been developped and studied by *Timmermans and von Sachs (2010, 2015)*, *Timmermans, Delsol and von Sachs (2013)* and *Timmermans and Fryzlewicz (2012), Fryzlewicz and Timmermans (2015)* and you should refer to these papers for detailed information about its use, and hence about the purpose of this package (see `http://hdl.handle.net/2078.1/154928` and `http://hdl.handle.net/2078.1/118369`, for curves and `http://hdl.handle.net/2078.1/110529` and `http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf` for images). Its approach is based upon the definition of a semi-distance that is functional, in the sense that the ordering of the data is explicitly taken into account, and wavelet-based, in the sense that it relies on a basis function expansion in which positions and amplitudes of a pattern are encoded. However, this method overcome the dyadic restriction that is attached with classical wavelet expansions, and do not require any preliminary smoothing of the data. Furthermore, a major originality of the method is that it relies on projections on basis functions that are different from one series to another.

Timmermans (2012) (http://hdl.handle.net/2078.1/112451) provides a comprehensive survey of the method. Regarding curves, the main ideas are as follows:

- Preliminary Observation. We consider series that are made of regularly spaced observations of a curve. We note that patterns in a series can be described as a set of level changes.
- Finding an *Optimal* Basis for Each Curve. As a first step, we want to expand each series in a basis that is best suited to it, in the sense that its first basis vectors should carry the main features of the series, while subsequent basis vectors support less significant patterns. In that respect, we are looking for a basis that is organized in a *hierarchical* way. As a consequence, there will be a particular basis associated to each series. As the series are thought of as described by their level changes, we will consider that the meaningful features for describing them are both locally important level changes (jumps, peaks, troughs) and level changes affecting a large number of data (discontinuities of the mean level). From this point of view, Unbalanced Haar wavelet bases are good candidates for our expansion. This family of bases have been introduced by *Girardi and Sweldens (1997)* as a way to circumvent the dyadic restriction of classical wavelets.
- Taking Advantage of the Hierarchy of those Bases. Given this, we will define a semi-distance, that is at the core of the BAGIDIS methodology. This semi-distance takes advantage of the hierarchy of the well-adapted unbalanced Haar wavelet bases: basis vectors of similar rank in the hierarchy (and their associated coefficients in the expansion of the series) are compared to each other, and the resulting differences are weighted according to that rank. This is actually a clue for decrypting the name of the methodology, as the name BAGIDIS stands for *BAsis GIving DIStances*.

Function `BUUHWE` computes the Unbalanced Haar Wavelet Expansion (BUUHWE) of a series in the Unbalanced Haar basis that is best suited to it according to the bottom-up algorithm developped by *Fryzlewicz (2007)*. Function `Breakpoints` and Function `Details` are used to encode the BUUHWE in a efficient compact way. Functions `BUUHWE.plot` and `BD.plot` give two different graphical representations of the BUUHWE expansion of a series. Function `BAGIDIS.dist` computes the BAGIDIS semidistance between two series encoded by their BUUHWE expansions. Function `BAGIDIS.dist.BD` computes the BAGIDIS semidistance between two series encoded by their breakpoints and details. `semimetric.BAGIDIS` computes the BAGIDIS semidistance between two series encoded as a numeric series. If `semimetric.BAGIDIS` is applied to matrices of series to be compared, it returns the dissimilarity matrix between them, as computed using the BAGIDIS semi-distance.

The BAGIDIS semi-distance between images relies on very similar principles. An optimal unbalanced Haar wavelet basis is associated to each image through the 2-dimensional Bottom-Up unbalanced Haar wavelet expansion (2D-BUUHWE). This expansion is rather termed SHape Adaptive Haar in *Timmermans and Fryzlewicz (2012), Fryzlewicz and Timmermans (2015)* and both terminology are used here in the functions name. The 2D-BUUHWE or SHAH is obtained thanks to function `BUUHWE_2D` and its alias `SHAH`. An efficient way to encode this expansion is through the *signature* of the image, obtained by applying `Signature_2D` to the output of `BUUHWE_2D` (or `SHAH`). A representation of the transform process is obtained using `BUUHWE_2D.plot` and its alias `SHAH.plot`. The transform process is linked with the definition of a well-suited unbalanced Haar basis for the image. This basis and its representation can be obtained using `BUUHWE_2D_Stepwise` and its alias SHAH_Stepwise. The BAGIDIS semi-distance between images is then computed using semimetric.BAGIDIS_2D.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain.
Maintainer: <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

The function BUUHWE_2D in this package is similar to the function uh.bu.2d (copyrighted Fryzlewicz 2014) in the package "shah_code", available on the webpage of Piotr Fryzlewicz: http://stats.lse.ac.uk/fryzlewicz/shah/shah_c
, which accompanies the paper Fryzlewicz and Timmermans (2015).

`BD.plot` *B-D representation of the Bottom Up Unbalanced Haar Wavelet Expansion of a series*

Description

Function for graphical representation of the Bottom Up Unbalanced Haar Wavelet Expansion (here after BUUHWE) of a series, in the *Breakpoints-Details* plane (B-D plane).

Usage

```
BD.plot(x, y = NULL, BUUHWE.out.x = BUUHWE(x), BUUHWE.out.y = BUUHWE(y),
French = FALSE, col = c('black','red'))
```

Arguments

`x`	a numeric vector (a series) whose BUUHWE expansion has to be computed and represented and the B-D plane
`y`	an optional second numeric vector (a series) whose `BUUHWE` expansion has to be computed and superimposed to the one of x in the B-D plane. `length(y)` must be equal to `length(x)`.
`BUUHWE.out.x`	output of `BUUHWE(x)`, included as an optional argument for saving computation time within the function if it has already been computed and saved outside the function; otherwise BUUHWE transform of x is compute at the call of function `BD.plot`.
`BUUHWE.out.y`	output of `BUUHWE(y)`, included as an optional argument for saving computation time within the function if it has already been computed and saved outside the function; otherwise the BUUHWE transform of y is computed by `BUUHWE` at the call of function `BD.plot`.
`French`	logical. Should labels be written in french ? (default=english)
`col`	vector of size one or two, indicating the colors for representing series x and - if needed -series y in the B-D plane. Default is black for series x and red for series y.

Details

See References below, in particular *Timmermans (2012)*, Chapter 1, or *Timmermans and von Sachs (2010)*.

Value

This function is invoked for its side effect which is to produce a graphical representation of the expansion of a series in its unbalanced Haar wavelet basis. The series is plotted in the plane that is defined by the values of its breakpoints and its detail coefficients. Points are numbered according to their rank in the hierarchy.

Note

Another way to represent the BUUHWE transform of a series is trough function BUUHWE.plot.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium.

Contact: <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

See Also

BUUHWE, BUUHWE.plot

Examples

```
x= c(1,7,3,0,-2,6,4,0,2)
BD.plot(x)
```

BUUHWE — *Bottom Up Unbalanced Haar Wavelet Expansion*

Description

Function of a time series x returning its unbalanced Haar wavelet expansion, as obtained with a bottom-up algortithm. BUUHWE is an acronym for Bottom-Up Unbalanced Haar Wavelets Expansion.

Usage

```
BUUHWE(x)
Breakpoints(Dataset.BUUHWE)
Details(Dataset.BUUHWE)
```

Arguments

`x`	a numeric vector of length N
`Dataset.BUUHWE`	a list of outputs of `BUUHWE` applied to a dataset of curves.

Details

See `BAGIDIS-package` for an overview about the BAGIDIS methodology and References for details, in particular *Timmermans (2012)*, Chapter 1, *Timmermans and von Sachs (2010)* or *Fryzlewicz (2007)*.

Value

`BUUHWE(x)` returns a list with

`detail`	detail coefficients starting from rank k=0 up to k=N-1.
`basis`	unbalanced Haar basis vectors ordered by colums. First column is the constant vector of rank k=0. This is a matrix of dimensions N xN.
`split.abs`	localization index. For consistency with basis matrix and detail vector dimensions, this is a vector of length N but first coefficient is `NA`. Index are then ordered by increasing rank k.
`series`	the initial series `x`.

`Breakpoints` returns the breakpoints of the Unbalanced Haar wavelet basis, from rank k=1 to rank k=N-1. In case it is applied to a list of BUUHWE provided for M series of length N, it returns the breakpoints for each elements of the list, as a matrix with M columns and N-1 rows.

`Details` returns the details of the Unbalanced Haar wavelet basis, from rank k=1 to rank k=N-1. In case it is applied to a list of outputs of `BUUHWE` provided for M series of length N, it returns the details for each elements of the list, as a matrix with M columns and N-1 rows.

Note

This function can be compared to function uh.bu in package **unbalhaar** of Piotr Fryzlewicz. Nevertheless outputs are differents as uh.bu is intented for denoising. The unbalanced Haar basis expansion is thus not explicitely obtained in uh.bu. Function BUUHWE.plot provides with a representation of the unbalanced Haar basis expansion.

Function BUUHWE_2D is the image equivalent of BUUHWE.

The BUUHWE expansion is the starting step for comparing curves through the BAGIDIS semi-distance. See function semimetric.BAGIDIS.

Breakpoints and details define what is called *the signature of the series in the b-d plane*. See references for details. A representation of the signature is obtained with BD.plot.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium.

Contact: <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

See Also

semimetric.BAGIDIS, BUUHWE.plot, BUUHWE_2D,BD.plot.

Examples

```
x= c(1,7,3,0,-2,6,4,0,2)
BUUHWE(x)
Breakpoints(list(BUUHWE(x)))
y= c(1,7,5,5,-2,1,4,0,2)
Breakpoints(list(BUUHWE(x),BUUHWE(y)))
```

BUUHWE.plot *Graphical representation of the Bottom Up Unbalanced Haar Wavelet Expansion of a series.*

Description

Function for graphical representation of the Bottom Up Unbalanced Haar Wavelet Expansion of a series, as obtained from function BUUHWE.

Usage

```
BUUHWE.plot(BUUHWE.out, row.max = length(BUUHWE.out$series),
French = FALSE, Color = TRUE)
```

Arguments

BUUHWE.out	output of function BUUHWE.
row.max	integer. Number of basis vectors to be represented. By default, all vectors are shown.
French	logical. Should labels be written in french? (default=english)
Color	logical. Should the representation be in color (alternative: black and white)? (default = color)

Details

See References below, in particular *Timmermans (2012)*, Chapter 1, or *Timmermans and von Sachs (2010)*.

Value

This function is invoked for its side effect which is to produce a graphical representation of the expansion of a series in its unbalanced Haar wavelet basis. The series is represented on the top of the graph, with the time axis at the bottom of the main figure. This main figure is the representation of the Unbalanced Haar basis, row by row, with rank k=0 (constant vector) on the top. Vectors are expressed as a function of time (horizontaly) and the height of that function is given by the *blue to red* color scale (if `Color = TRUE`) or relatively to a zero level, that is indicated by a dotted horizontal line for each rank. Vertically, the detail coefficients are represented rank by rank at the same level as the corresponding basis vectors. This is the expansion of the series in its associated Unbalanced Haar basis, starting from rang k= 0 on the top of the gaph - wich is represented missing for graphical purpose and continuing downwards with rang k increasing. If you sum those detail coefficients times their associated basis vectors, you get back the initial series.

Note

The equivalent of `BUUHWE.plot` for images is `BUUHWE_2D.plot`. Another way to represent the BUUHWE transform of a series is through function `BD.plot`.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium.

Contact: `<catherine.timmermans@uclouvain.be>`

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.

- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

The function BUUHWE_2D in this package is similar to the function uh.bu.2d (copyrighted Fryzlewicz 2014) in the package "shah_code", available on the webpage of Piotr Fryzlewicz: http://stats.lse.ac.uk/fryzlewicz/shah/shah_c
, which accompanies the paper Fryzlewicz and Timmermans (2015).

See Also

BUUHWE, BD.plot,BUUHWE_2D.plot

Examples

```
x = c(0,0,1,0,0, 3, 5, 0, 0)
BUUHWE.plot(BUUHWE(x))
BUUHWE.plot(BUUHWE(x), Color= FALSE)
```

BUUHWE_2D   *Two-dimensional Bottom-Up Unbalanced Haar Wavelet Expansion of an image*

Description

This function computes the two-dimensional Bottom-Up Unbalanced Haar Wavelet Expansion (2D-BUUHWE) of an image (encoded as a matrix). See Timmermans (2012) for details. The 2D-BUUHWE is termed SHaped-Adaptive Haar (SHAH) expansion in Timmermans and Fryzlewicz (2012).

Usage

```
BUUHWE_2D(im)
SHAH(im)
Signature_2D(out.BUUHWE_2D)
```

Arguments

im	im: a numeric matrix encoding an image. Missing values are not allowed.
out.BUUHWE_2D	out.BUUHWE_2D: the output of BUUHWE_2D or SHAH.

Details

More details can be found in Timmermans (2012), Chapter 4, or Timmermans and Fryzlewicz (2012). The function implements the 2D-BUUHWE (= SHAH) for images observed on a regular quadratic grid. The image is encoded as a matrix, of which the entries are labeled by column.

Value

The output of BUUHWE_2D or SHAH is a list with

d
: a vector of two elements giving the dimension of the encoded image.

decomp.hist
: The decomposition history of the image, which provides its BUUHWE expansion (=SHAH transform). For each i in d[1]*d[2]-1, matrix decomp.hist[,,i] is a matrix with the following entries: decomp.hist[1,1,i] is the label of the input node at step d[1]*d[2]-i, decomp.hist[1,2,i] is the label of the output node at step d[1]*d[2]-i, decomp.hist[2,1,i] is the weight of the input node at step d[1]*d[2]-i, decomp.hist[2,2,i] is the weight of the output node at step d[1]*d[2]-i , decomp.hist[3,1,i] is the detail d[1]*d[2]-i, decomp.hist[3,2,i] is always zero.

The output of Signature_2D is a matrix containing the signature of the image. This is a matrix of which row i is made of the components x1,y1,x2,y2, detail and abs(detail) related to rank i in the 2D-BUUHWE (SHAH) of an image.

Note

A graphical representation of the tranform is provided by BUUHWE_2D.plot (or equivalently SHAH.plot). The 2D-BUUHWE (SHAH) is related to a basis expansion. The basis can be obtained and represented using BUUHWE_2D_Stepwise (or equivalently SHAH_Stepwise). An efficient way to store the information defing the 2D-BUUHWE (SHAH) is through the signature of the image, obtained using Signature_2D.

BUUHWE provides with the Bottom-Up Unbalanced Haar Wavelet Expansion of a curve. Pairs of images can be compared through their 2D-BUUHWE, using the BAGIDIS semi-distance computed with semimetric.BAGIDIS_2D. See Timmermans (2012) for details.

Author(s)

Piotr Fryzlewicz, Department of Statistics, London School of Economics.

Catherine Timmermans, Institut de Statistique, Biostatistique et Sciences Actuarielles, Universite catholique de Louvain. <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf

- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

The function BUUHWE_2D in this package is similar to the function uh.bu.2d (copyrighted Fryzlewicz 2014) in the package "shah_code", available on the webpage of Piotr Fryzlewicz: http://stats.lse.ac.uk/fryzlewicz/shah/shah_c
, which accompanies the paper Fryzlewicz and Timmermans (2015).

See Also

BUUHWE_2D.plot,SHAH.plot ,BUUHWE_2D_Stepwise, SHAH_Stepwise, BUUHWE.

Examples

```
im = rbind(c(1,2,3),c(2,7,8),c(1,0,0))
BUUHWE_2D(im)
SHAH(im)
Signature_2D(SHAH(im))

SHAH_Stepwise(im)$details
Signature_2D(SHAH(im))
```

BUUHWE_2D.plot *Graphical representation of the 2D-BUUHWE (or SHAH) of an image*

Description

Graphical representation of the 2D-BUUHWE or SHAH transform of an image. The representation of the process is either through the evolving network representation or through the successively built basis matrices of the Unbalanced Haar wavelet basis associated to the transform.

Usage

```
BUUHWE_2D.plot(out.BUUHWE_2D, Evol = TRUE)
SHAH.plot(out.BUUHWE_2D, Evol = TRUE)
BUUHWE_2D_Stepwise(Data,Plot)
SHAH_Stepwise(Data,Plot)
```

Arguments

out.BUUHWE_2D	the result of BUUHWE_2D or SHAH applied to an image.
Evol	Logical. Should the evolution of the complete network be represented? Default is TRUE and is more informative but applies only for images of reasonable size (otherwise graphics are not readable). if FALSE, only segments linking the input and output nodes at each step are represented and superimposed.
Data	a matrix encoding an image
Plot	Logical. Should all the basis matrices be represented? Default is TRUE. Otherwise not representation is provided, but the matrices are returned by the function so that they can be represented separetely.

Details

See BAGIDIS-package for an overview about the BAGIDIS methodology and References for details, in particular *Timmermans (2012)*, Chapter 4, and *Timmermans and Fryzlewicz (2012)*.

Value

BUUHWT_2D_Stepwise or SHAH_Stepwise returns a list with

basis	An array of the basis matrices of the 2D-BUUHWT (SHAH)
details	the detail coefficients for ranks NM-1 to 1 (with N*M the dimension of the image) (i.e. in their ordre of construction)
d0	the detail coefficients for ranks 0
im	the initial image
labels.hist	the evolution of the labels associated to the pixels

The function are mainly used for their side-effect Plot=TRUE. Functions BUUHWE_2D.plot or SHAH.plot are only used for their side-effect.

Note

The equivalent function for curves is semimetric.BAGIDIS.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium.

Contact: <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451

- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

See Also

BUUHWE_2D, SHAH, semimetric.BAGIDIS.

Examples

```
im = rbind(c(1,2,3), c(5,3,2), c(1,1,2))
im2 = rbind(c(1,1,5), c(5,5,2), c(1,0,0))

SHAH(im)
SHAH_Stepwise(im)

SHAH.plot(SHAH(im))

SHAH.plot(SHAH(im),Evol=FALSE)

SHAH_Stepwise(im)$details
Signature_2D(SHAH(im))
```

`semimetric.BAGIDIS` *Computing the BAGIDIS semidistance between series*

Description

Functions for computing the Bagidis semidistance between series of measurements.

Usage

```
semimetric.BAGIDIS(DATA1,DATA2=DATA1, p = 2,  wk=NULL, Param=0.5,
 wdw= min(ncol(DATA1),30),   Evol =FALSE,  Overlap = wdw-1,  method = c('TS','BD'))

semimetric.BAGIDIS.TS(DATA1,DATA2=DATA1, p = 2,  wk=NULL, Param=0.5,
 wdw= min(ncol(DATA1),30),   Evol =FALSE,  Overlap = wdw-1)

semimetric.BAGIDIS.BD(Details1, Breakpoints1,Details2=Details1,
 Breakpoints2=Breakpoints1,NbSubseries =1,p = 2,  wk=NULL, Param=0.5, Evol =FALSE)

BAGIDIS.dist(BUUHWE.out.1, BUUHWE.out.2, p = 2, wk=NULL, Param=0.5)

BAGIDIS.dist.BD(Details1, Breakpoints1,Details2, Breakpoints2,
p = 2, wk=NULL, Param=0.5)
```

Arguments

DATA1	matrice containing the series to be compared row by row.with the rows of `DATA2` Each row of `DATA1` has its semidistance being computed with every row of `DATA2`. If only `DATA1` is provided, then `DATA2` is taken to be equal to `DATA1`. We must have `ncol(DATA1)= ncol(DATA2)`.
DATA2	Optional. Matrice containing the series to be compared row by row.with the rows of `DATA1` Each row of `DATA1` has its semidistance being computed with every row of `DATA2`. If only `DATA1` is provided, then `DATA2` is taken to be equal to `DATA1`. We must have `ncol(DATA1)= ncol(DATA2)`.
p	the kind of norm to be used for computing the partial distance in the B-D plane. Must be numeric or `Inf`.
wk	a vector of weights of length `ncol(DATA1)-1`. If not provided, `wk = log(N+1-(1:N))/log(N+1)` with `N= ncol(DATA1)-1`.
Param	the balance parameter between the differences along the breakpoint axis and along the detail axis. Param must be in [0;1]. `Param= 1` means that only breakpoints differences are taken into account. `Param=0` means that only details differences are taken into account.
wdw	In case distances are measured between "long" series, it could be advantageous to make use of a windowed semimetric. `wdw` encode the length of the window in which the semimetric will be computed between the subseries. By default there is no windowing if the length of the series (`= ncol(DATA1)`) is smaller than or equal to 30, and a windows length of 30 otherwise.

Evol	Logical. In case a windowing is applied, should the matrices of local (windowed) dissimilarities be returned? Default is FALSE.
Overlap	In case a windowing of the series is applied, Overlap determines how the subseries overlap each other. By default, a one-step-sliding distance is computed.
method	either 'TS' (default) or 'BD' : the method for computing the matrix of semi-distances in case of multiple series. Results are identical. 'TS' recompute the BUUHWE transform for each pairwise comparison, 'BD' computes all signatures beforehand and store them before computing the distances. 'TS' requires more time, 'BD' requires more storage. With method 'TS', computation time is affected by the number of rows in DATA1 and DATA2. If nrow(DATA1)> nrow(DATA2), it increases the number of operations to be computed. On the opposite, if nrow(DATA2)>nrow(DATA1), it increases the memory usage.
BUUHWE.out.1	BUUHWE expansion of a series, as obtained from function BUUHWE.
BUUHWE.out.2	BUUHWE expansion of a second series, as obtained from function BUUHWE.
Details1	matrixcontaining the details of series out of a dataset DATA1 containing a set of series of identical length.
Breakpoints1	matrix containing the breakpoints of series out of a dataset DATA1 containing a set of series of identical length.
Details2	matrixcontaining the details of series out of a dataset DATA2 containing a set of series of identical length as in DATA1.
Breakpoints2	matrix containing the breakpoints of series out of a dataset DATA2 containing a set of series of identical length as in DATA1.
NbSubseries	in case an evolving (windowed) semidistance must be computed, Nbsubseries gives the number of data measurements in a windowed segment.

Details

Function semimetric.BAGIDIS computes the Bagidis semidistance between curves. If several curves are provided, it returns a matrix of semidistances. The function is an interface for either semimetric.BAGIDIS.TS or semimetric.BAGIDIS.BD, depending on the value of the parameter method.

Function Bagidis.dist computes the BAGIDIS semidistance between two series, encoded through their BUUHWE expansion obtained from function BUUHWE. Function Bagidis.dist.BD computes the BAGIDIS semidistance between two series, encoded through their breakpoints and details obtained from functions Breakpoints and Details.

See BAGIDIS-package for an overview about the BAGIDIS methodology and References for details, in particular *Timmermans (2012)*, Chapter 1, and *Timmermans and von Sachs (2010)*.

Value

- dissimilarity.matrix Matrix of semidistances between the nrow(DATA1) series of DATA1 and the nrow(DATA2) series of DATA2. Dimensions: nrow(DATA1) x nrow(DATA2) .
- dissimilarity.evol Array of local matrices of semidistances between the windowed series of DATA1 and DATA2. Dimensions: nrow(DATA1) x nrow(DATA2) x Nb_subseries. Nb_subseries is determined by the three quantities nrow(DATA1), wdw and Overlap.

Note

The equivalent function for image is semimetric.BAGIDIS_2D.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium.

Contact: <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

The function BUUHWE_2D in this package is similar to the function uh.bu.2d (copyrighted Fryzlewicz 2014) in the package "shah_code", available on the webpage of Piotr Fryzlewicz: http://stats.lse.ac.uk/fryzlewicz/shah/shah_c
, which accompanies the paper Fryzlewicz and Timmermans (2015).

See Also

BUUHWE, semimetric.BAGIDIS_2D.

Examples

```
x= 1:10
y=2:11
A=rbind(x,y)
semimetric.BAGIDIS(A)

B= rbind(x,x,y)
semimetric.BAGIDIS(A,B)

x= 1:30
y= 1:30
A= rbind(x,y)
B= rbind(x,x, y)
semimetric.BAGIDIS(A,B, wdw =15, Evol =TRUE, Overlap =0)

x= c(1,7,3,0,-2,6,4,0,2)
y= c(1,7,5,5,-2,1,4,0,2)
BAGIDIS.dist(BUUHWE(x), BUUHWE(y))
```

`semimetric.BAGIDIS_2D` *The BAGIDIS semi-distance for datasets of images*

Description

Function for computing the Bagidis semidistance between images.

Usage

```
semimetric.BAGIDIS_2D(Data1, Data2 = Data1, NbROW = NULL, wk=NULL,
 lambdaV = NULL, lambdaH = NULL, Type = c( "Array","Vector", "BD", "Flex"))
```

Arguments

`Data1`	A dataset of images. The way this dataset is encoded depend on Type. See Details. There must be at least two images.
`Data2`	A dataset of images encoded in the same way as `Data1`.
`NbROW`	If `Type` = "Vector", the number of row of the matrix representation associated to the image. Otherwise ignored.
`wk`	The weight function. For images of size N*M, this is a vector of length N*M-1 if `Type` in c("Vector", "Array", "BD") and a matrix of N*M-1 row and 6 columns if `Type` ="Flex". In the latter case, the columns refer to the weights associated to x1,y1,x2,y2, details and abs(details) respectively and the rows refer to the rank k = 1 to N-1 in the 2D-BUUHWE (SHAH). The 6th columns is usually 0. See References for details.

lambdaV — If Type in c("Vector", "Array", "BD"), the scaling parameter in the vertical direction. Otherwise ignored. It must be in [0,1] with lambdaV+ lambdaH <=1.

lambdaH — If Type in c("Vector", "Array", "BD"), the scaling parameter in the vertical direction. Otherwise ignored. It must be in [0,1] with lambdaV+ lambdaH <=1.

Type — One amongst ("Array","Vector" "BD","Flex"). This indicates the way Data1 and Data2 encode the images to be compared. It also accounts for the specific form used for the BAGIDIS semidistance. See Details below.

Details

if Type="Vector" , the images are stored as vector of values encoded by row in the matrix Data1 (and Data2. There is one row per image and one column per pixel. The parameter NROW is required to specify the shape of the image (which is then a matrix constructed row by row).

If Type= "Array" , the images are stored as matrices along the third dimension of the array Data1 (or Data2). For instance, the first image is Data1[,,1].

If Type= "BD" or "Flex" , the images are encoded through their signatures (obtained from Signature_2D). Data1 (or Data2) is a list of signatures.

If Type in c("Vector", "Array", "BD"), the BAGIDIS semi-distance is used in its constrained form, with a vector weight function and three scaling parameters (lambdaV along the vertical direction, lambdaH along the horizontal diirection and a scaling parameter along the detail direction) which sum to 1. See References for more details, in particular *Timmermans (2012)*, Chapter 4, and *Timmermans and Fryzlewicz (2012)*. The values of lambdaV and lambdaH fix the scaling in the detail direction as 1 - lambdaV - lambdaH, which must remain nonnegative. Default values are provided if wk, lambdaV and lambdaH are nonspecified, as follows: wk = log(N+1-(1:N))/log(N+1) with N the number of pixels - 1, lambdaV=lambdaH=1/3 if both are nonspecified, lambdaH= (1-lambdaV)/2 or lambdaV= (1-lambdaH)/2 if only one is supplied.

If Type ="Flex", the BAGIDIS semi-distance is used in its general form with the matrix wk encoding a matrix weighting specifying the components of the signature at each rank. See References below, in particular *Timmermans (2012)*, Chapter 4, and *Timmermans and Fryzlewicz (2012)*. The absolute value of the detail is included in the signature, but can be ignored by defining the 6th column of wk as a vector of 0.

If only two images have to be compared to each other, both must be in Data1, and Data2 = Data1.

Value

The matrix of Bagidis semidistances between the nrow(DATA1) ima ges of DATA1 and the nrow(DATA2) images of DATA2. Dimensions: nrow(DATA1) x nrow(DATA2) .

Note

The equivalent function for curves is semimetric.BAGIDIS.

Author(s)

Catherine Timmermans, Institute of Statistics, Biostatistics and Actuarial Sciences, UCLouvain, Belgium.

Contact: <catherine.timmermans@uclouvain.be>

References

The main references are

- Timmermans C., 2012, Bases Giving Distances. A new paradigm for investigating functional data with applications for spectroscopy. PhD thesis, Universite catholique de Louvain. http://hdl.handle.net/2078.1/112451
- Timmermans C. and von Sachs R., 2015, A novel semi-distance for investigating dissimilarities of curves with sharp local patterns, Journal of Statistical Planning and Inference, 160, 35-50. http://hdl.handle.net/2078.1/154928
- Fryzlewicz P. and Timmermans C., 2015, SHAH: Shape Adaptive Haar wavelets for image processing. Journal of Computational and Graphical Statistics. (accepted - published online 27 May 2015) http://stats.lse.ac.uk/fryzlewicz/shah/shah.pdf
- Timmermans C., Delsol L. and von Sachs R., 2013, Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features, Journal of Multivariate Analysis, 115, p. 421-444. http://hdl.handle.net/2078.1/118369

Other references include

- Girardi M. and Sweldens W., 1997, *A new class of unbalanced Haar wavelets that form an unconditional basis for Lp on general measure spaces*, J. Fourier Anal. Appl. 3, 457-474
- Fryzlewicz P., 2007, *Unbalanced Haar Technique for Non Parametric Function Estimation*, Journal of the American Statistical Association, 102, 1318-1327.
- Timmermans C., von Sachs, R. , 2010, *BAGIDIS, a new method for statistical analysis of differences between curves with sharp patterns* (ISBA Discussion Paper 2010/30). Url : http://hdl.handle.net/2078.1/91090
- Timmermans, C. , Fryzlewicz, P., 2012, *SHAH: Shape-Adaptive Haar Wavelet Transform For Images With Application To Classification* (ISBA Discussion Paper 2012/15). Url: http://hdl.handle.net/2078.1/110529

See Also

BUUHWE_2D, SHAH, semimetric.BAGIDIS.

Examples

```
im = rbind(c(1,2,3), c(5,3,2), c(1,1,2))
im2 = rbind(c(1,1,5), c(5,5,2), c(1,0,0))

Data1= rbind(as.numeric(im),as.numeric(im2))
semimetric.BAGIDIS_2D(Data1, NbROW= 3, Type= 'Vector')

Data1= abind(im,im2, along=3)
semimetric.BAGIDIS_2D(Data1, Type= 'Array')

Data1= list(Signature_2D(SHAH(im)),Signature_2D(SHAH(im2)))
semimetric.BAGIDIS_2D(Data1, Type= 'BD')

Data1= list(Signature_2D(SHAH(im)),Signature_2D(SHAH(im2)))
wk = matrix(0, nrow=nrow(Data1[[1]]), ncol=ncol(Data1[[1]]))
```

```
wk[1:2,1:5] = rep(1,10)
semimetric.BAGIDIS_2D(Data1, Type= 'Flex', wk=wk)
```

Index