Package 'changepoint.influence'

August 4, 2021

Type Package
Title Package to Calculate the Influence of the Data on a Changepoint Segmentation
Version 1.0.1
Date 2021-08-03
Maintainer Rebecca Killick < r. killick@lancs.ac.uk>
<pre>BugReports https://github.com/rkillick/changepoint.influence/issues</pre>
<pre>URL https://github.com/rkillick/changepoint.influence/</pre>
Imports data.table, ggplot2, gridExtra, reshape, graphics, methods
Depends $R(>=3.2)$, changepoint
Suggests testthat, vdiffr
Description Allows users to input their data, segmentation and function used for the segmentation (and additional arguments) and the package calculates the influence of the data on the changepoint locations, see Wilms et al. (2021) <arxiv:2107.10572>. Currently this can only be used with the changepoint package functions to identify changes, but we plan to extend this. There are options for different types of graphics to assess the influence.</arxiv:2107.10572>
License GPL
LazyData true
NeedsCompilation no
Author Rebecca Killick [aut, cre], Ines Wilms [aut]
Repository CRAN
Date/Publication 2021-08-04 15:00:02 UTC
R topics documented:
changepoint.influence-package 2 InfluenceMap 4 LocationStability 7

ParameterStability									 								9
StabilityOverview									 								10
welldata									 								12

Index 14

changepoint.influence-package

Package to Calculate the Influence of the Data on a Changepoint Segmentation

Description

Allows users to input their data, segmentation and function used for the segmentation (and additional arguments) and the package calculates the influence of the data on the changepoint locations, see Wilms et al. (2021) <arXiv:2107.10572>. Currently this can only be used with the changepoint package functions to identify changes, but we plan to extend this. There are options for different types of graphics to assess the influence.

Details

The DESCRIPTION file:

Package: changepoint.influence

Type: Package

Title: Package to Calculate the Influence of the Data on a Changepoint Segmentation

Version: 1.0.1 Date: 2021-08-03

Authors@R: c(person("Rebecca", "Killick", role=c("aut", "cre"), email="r.killick@lancs.ac.uk"), person("Ines", "Wilms", role=c("aut", "cre"), email="r.killick@lancs.ac.uk"), person("aut", "cre"), email="r.killick@lancs.ac.uk"), person("aut", "cre"), email="r.killick@lancs.ac.uk"), person("aut", "cre"), email="r.killick@lancs.ac.uk"), email="r.k

Maintainer: Rebecca Killick <r.killick@lancs.ac.uk>

BugReports: https://github.com/rkillick/changepoint.influence/issues URL: https://github.com/rkillick/changepoint.influence/

Imports: data.table, ggplot2, gridExtra, reshape, graphics, methods

Depends: R(>= 3.2), changepoint

Suggests: testthat, vdiffr

Description: Allows users to input their data, segmentation and function used for the segmentation (and additional argumen

License: GPL LazyData: true

Packaged: 2021-06-22 15:55:29 UTC; killick

Author: Rebecca Killick [aut, cre], Ines Wilms [aut]

Index of help topics:

InfluenceMap Influence Map Graphic
LocationStability Location Stability Graphic
ParameterStability Parameter Stability Graphic
StabilityOverview Stability Overview Graphic

changepoint.influence-package

Package to Calculate the Influence of the Data

on a Changepoint Segmentation

welldata Welllog data

The package allows users to input their data, segmentation and function used for the segmentation (and additional arguments) and the package calculates the influence of the data on the changepoint locations.

The influence() function is the first port of call to calculate the influence. We provide two methods for influence detection, via "delete" and "outlier" options which respectively consider the effect of deleting a data point or making it an outlier. Currently we provide this method for cpt objects (as generated by the "changepoint" package) but plan to extend this to other objects in the future. Please add requests for objects to include to our github issues.

Users are encouraged to explore the documentation for the StabilityOverview() graphic, followed by the LocationStability() and ParameterStability() graphics for a more granual view, followed by the InfluenceMap() as the highest level of detail.

Author(s)

NA

Maintainer: Rebecca Killick <r.killick@lancs.ac.uk>

References

Wilms I, Killick R, Matteson DS (2021) Graphical Influence Diagnostics for Changepoint Models, *ArxiV*: 2107.10572

See Also

StabilityOverview, LocationStability, ParameterStability, InfluenceMap

Examples

```
#### Load the data in the R package changepoint.influence ####
data("welldata")
welllog = welldata[1001:2000] # Extract the mid section of the data as analyzed in other papers
n = length(welllog)
var = NULL; for (i in 30:1000){var[i]=var(welllog[(i-29):i])}
welllogs = welllog/sqrt(median(var, na.rm = TRUE))
# rescale the data to have unit variance across time,
# note that there may still be changes in variance across the series.

#### Apply PELT to the welllog data ####
out.PELT = cpt.mean(welllogs, method = 'PELT')

#### Calculate the influence measures ####
welllogs.inf = influence(out.PELT)
# the code extracts all the details of the original cpt.mean() function call
# and uses these in the calculation of the influence for the modified data.
```

4 InfluenceMap

```
#### Stability Dashboards ####
StabilityOverview(welllogs, cpts(out.PELT), welllogs.inf, las = 1,ylab='Nuclear-Magnetic Response',
 legend.args=list(display=TRUE,x="bottomright",y=NULL,cex=1.5,bty="n",horiz=FALSE,xpd=FALSE))
# We can specify where the legend will sit in the graphic via the legend.args
# which are passed to the legend() function. We can also include additional arguments
# to pass to the plotting such as las=1 here.
#### Location Stability plot ####
exp.seg=LocationStability(cpts(out.PELT), welllogs.inf, type = 'Difference', cpt.lwd = 4, las = 1)
# Note that if the expected segmentation is not provided, it will be calcuated and then
# returned so that the user can avoid calculating this again in other plot calls.
#### Parameter Stability plot ####
ParameterStability(welllogs.inf, original.mean = rep(param.est(out.PELT)$mean,
  times=diff(c(0,out.PELT@cpts))), las = 1, ylab = 'Nuclear-Magnetic Response')
# Note that the original.mean argument is provided for each timepoint so is a length n vector.
#### Influence Map ####
## Not run:
library(ggplot2)
welllogs.inf = influence(out.PELT, method = "delete")
InfluenceMap(cpts(out.PELT), welllogs.inf, data=welllogs, include.data=TRUE,
   ylab='Nuclear-Magnetic\n Response',
    ggops=theme(axis.text=element_text(size=15),axis.title=element_text(size=20),
      plot.title=element_text(size=25)))
# The InfluenceMap uses ggplot2 functions, thus you can add theme options via the ggops argument.
# Here we change the text sizes to ensure readable titles and labels for a report.
welllogs.inf = influence(out.PELT, method = "outlier")
InfluenceMap(cpts(out.PELT), welllogs.inf, data = welllogs, include.data = TRUE,
   ylab='Nuclear-Magnetic\n Response')
## End(Not run)
```

InfluenceMap

Influence Map Graphic

Description

Plots the highest detail level of the changepoint location stability according to the influence measure.

Usage

```
InfluenceMap(original.cpts, influence, resid=NULL,data=NULL,include.data=FALSE,
influence.col=c("#0C4479","white","#AB9783"),cpt.col=c("#009E73", "#E69F00", "#E41A1C"),
cpt.lty=c("dashed","dotdash","dotted"),ylab='',ggops=NULL)
```

InfluenceMap 5

Arguments

original.cpts	An ordered vector of the changepoint locations found by your favourite changepoint method.
influence	The influence as calculated the influence() function provided within this package. This is a list object.
resid	An nxn matrix containing the difference of the observed class (influence) from the expected class at each datapoint. If this is left as NULL, it will be calculated and returned to the user.
data	A vector containing the data on which you have run your changepoint method.
include.data	Is a plot of the data to be included above the histogram. Default is FALSE.
influence.col	A length 3 vector giving the lower, middle (0) and upper bounds for the influence map colour grading. Note that you should choose these colours to not conflict with the colours used for cpt.col. We advise using "white" for the middle choice as this provides a clean (majority white) heatmap.
cpt.col	Colour of the original.cpts lines when plotted. We need three colours specified here, corresponding to the "stable", "unstable" and "outlier" categories respectively. These are plotted on the Influence Map as well as the original data if include.data=TRUE. Any values accepted by the col plotting argument are allowed.
cpt.lty	Line type of the original.cpts lines when plotted. We need three line types specified here, corresponding to the "stable", "unstable" and "outlier" categories respectively. Any values accepted by the 1ty plotting argument are allowed. Only used if include.data=TRUE.
ylab	The label for the y-axis, character vector expected.
ggops	Any other settings to be passed to the ggplot() function. Note that you will need to library(ggplot2) if calling ggplot functions in arguments. See examples.

Details

This function creates the highest detail graphic to display the results of a changepoint influence analysis on the location of the changepoints. The graphic is an nxn heatmap of the difference between the observed segmentations under the "delete" or "outlier" Influence analysis and the expected segmentation. Note that the expected segmentations take into account the fact that a changepoint at a timepoint, say 100, will move (to 99) when a timepoint prior to it is deleted and that adding an outlier will introduce new changepoints.

Datapoints on the vertical axis without a single coloured co-ordinate on the horizontal axis can be considered as non-influential since they do not trigger any changepoint instability. Rows with coloured pixels correspond to data points which are instability triggers.

How to interpret the Influence Map (please also read the paper in the references for fuller details):

• colouring:Colouring above the diagonal indicates that an al-teration of the corresponding data point (on the vertical axis) affects earlier data points, colouring below the diagonal indicates that subsequent data points are affected.

6 InfluenceMap

 horizonal span: A stop in colouring indicates that change-points have moved, while a continuation of colouring to the last data point indicates that, in total, fewer or additional changepoints are detected.

- local vs global:Most colouring originates on the diagonal,thereby indicating that a data point's alteration mainly affects neighbouring data points that most often belong to the same segment. By contrast, in some cases a coloured pixel may originate away from the diagonal, thereby exercising global influence.
- height:All data points (on the vertical axis) that appear in the coloured area are influential and assert influence over the corresponding data points on the horizontal axis. The height can be seen as the extent to which instability arises in this influential region.

Value

The function returns a plot denoted the Influence Map. If resid=NULL then the residuals (observed class - expected class) are also returned.

Author(s)

Rebecca Killick

References

Wilms I, Killick R, Matteson DS (2021) Graphical Influence Diagnostics for Changepoint Models, *ArxiV*: 2107.10572

See Also

```
influence, StabilityOverview, ParameterStability, LocationStability
```

Examples

LocationStability 7

End(Not run)

LocationStability Location Stability Graphic

Description

Plots the middle detail level of the changepoint location stability according to the influence measure.

Usage

```
LocationStability(original.cpts, influence, expected.class=NULL,
   type=c("Difference","Global","Local"),data=NULL,include.data=FALSE,cpt.lwd=4,
   cpt.col=c("#009E73", "#E69F00", "#E41A1C"),cpt.lty=c("dashed","dotdash","dotted"),
   ylab='',xlab='Index',...)
```

Arguments

original.cpts	An ordered vector of the changepoint locations found by your favourite change- point method.
influence	The influence as calculated the influence() function provided within this package. This is a list object.
expected.class	Only needed for type="Difference". An nxn matrix containing the expected class of each datapoint from the original.cpts segmentation under perturbation of each datapoint. If this is left as NULL and type="Difference", it will be calculated and returned to the user.
type	The type of Location Stability plot, can be "Difference", "Global" or "Local". If all three are listed (as in the default) then three graphs will be plotted. type="Difference" will histogram the difference between the observed and expected segmentations at each of the datapoints. type="Global" will histogram the observed segmentations. type="Local" will histogram the observed segmentations with the original.cpts locations removed. See details for more description of the graphics.
data	A vector containing the data on which you have run your changepoint method.
include.data	Is a plot of the data to be included above the histogram. Default is FALSE.
cpt.lwd	The line width to be used when plotting the original.cpts on the data. Standard lwd values allowed. A single value is expected.
cpt.col	Colour of the original.cpts lines when plotted. We need three colours specified here, corresponding to the "stable", "unstable" and "outlier" categories respectively. Any values accepted by the col plotting argument are allowed.
cpt.lty	Line type of the original.cpts lines when plotted. We need three line types specified here, corresponding to the "stable", "unstable" and "outlier" categories respectively. Any values accepted by the 1ty plotting argument are allowed.
ylab, xlab	The labels for the x- and y-axis, character vector expected.
	Any other arguments to be passed to the plot() or equivalently hist() function.

8 LocationStability

Details

This function creates a more granular graphic to display the results of a changepoint influence analysis on the location of the changepoints. The graphic is a histogram of the observed segmentations under the "delete" or "outlier" Influence analysis. The colour and line type of the bars at the original.cpts locations reflect their stability. The first value of their arguments denotes a stable changepoint - which appears at the same location in all influence segmentations. The second argument denotes an unstable changepoint - which doesn't appear at the same location in all influence segmentations, either it moves or is deleted. The third argument denotes changepoint locations which are deemed outliers as two changepoints occur at consecutive locations (surrounding the outlying observation). Please note that the type="Global" only uses colour and not line type.

type="Difference" gives the difference between the observed and expected changepoint segmentations under the "delete" or "outlier" Influence analysis. A positive value can only occur where a changepoint is contained in the observed segmentations but is not present in the expected (an additional changepoint time). A negative value can only occur at the original changepoint location where the changepoint is not present in atleast one of the observed segmentations. Note that the expected segmentations take into account the fact that a changepoint at a timepoint, say 100, will move (to 99) when a timepoint prior to it is deleted.

type="Global" histograms the observed segmentations. Colour is added to the original change-point locations and a horizontal (light grey) line is added to the plot to denote the maximum count. Any original changepoint bars that do not meet this grey line indicates that the changepoint is unstable as it either moves or is deleted in atleast one of the observed segmentations. For large datasets this can be difficult to view what is going on at any locations that appear as black bars as these are typically small counts. Hence the inclusion of the "Local" option.

type="Local" histograms the observed segmentations with the original changepoint locations removed. This is to allow users to see the smaller counts that can be masked in larger datasets. These are the locations where either original changepoints move to or additional changepoints are added.

Value

The function returns plot(s) and a list containing the labels of the original.cpts as either "stable", "unstable", or "outlier". If type="Difference" and expected.class=NULL then the expected class is also returned as the first element of the list.

Author(s)

Rebecca Killick

References

Wilms I, Killick R, Matteson DS (2021) Graphical Influence Diagnostics for Changepoint Models, *ArxiV*: 2107.10572

See Also

influence, StabilityOverview,ParameterStability,InfluenceMap

ParameterStability 9

Examples

```
#### Generate Simulated data example ####
set.seed(30)
x = c(rnorm(50), rnorm(50, mean = 5), rnorm(1, mean = 15), rnorm(49, mean = 5), rnorm(50, mean = 4))
xcpt = cpt.mean(x,method='PELT') # Get the changepoints via PELT

#### Get the influence for both delete and outlier options ####
x.inf = influence(xcpt)

#### Location Stability Difference plot ####
exp.class=LocationStability(cpts(xcpt), type = 'Difference', x.inf, cpt.lwd = 4, las = 1)
# note that the expected.class is also returned

#### Location Stability Global plot ####
exp.class=LocationStability(cpts(xcpt), type = 'Global', x.inf, cpt.lwd = 4, las = 1)

#### Location Stability Local plot ####
exp.class=LocationStability(cpts(xcpt), type = 'Local', x.inf, cpt.lwd = 4, las = 1)
```

ParameterStability

Parameter Stability Graphic

Description

Plots the middle detail level of the changepoint parameter stability according to the influence measure.

Usage

```
ParameterStability(influence,original.mean=NULL,digits=6,ylab='',xlab='Index',
    cpt.col='red',cpt.width=3,...)
```

Arguments

influence	The influence as calculated the influence() function provided within this package. This is a list object.
original.mean	A vector, $length n$, of the mean under the original segmentation at each time-point.
digits	The number of significant figures to round the mean values to before plotting. (Purely to reduce the number of points plotted to make the graphics smaller for storage and loading)
ylab, xlab	The labels for the x- and y-axis, character vector expected.
cpt.col	Colour of the original parameter vector when plotted. Any values accepted by the col plotting argument are allowed.
cpt.width	Width of the original parameter vector when plotted. Any values accepted by the 1wd plotting argument are allowed.
	Any other arguments to be passed to the plot() function.

10 StabilityOverview

Details

This function creates a more granular graphic to display the results of a changepoint influence analysis on the estimated segment parameter. The graphic depicts the observed segment parameters under the "delete" or "outlier" Influence analysis. The intensity of the grey denotes how often that parameter values was seen across all segmentations. We overlay this with the original segment parameters.

Value

The function returns a plot (silently).

Author(s)

Rebecca Killick

References

Wilms I, Killick R, Matteson DS (2021) Graphical Influence Diagnostics for Changepoint Models, *ArxiV*: 2107.10572

See Also

influence, StabilityOverview,LocationStability,InfluenceMap

Examples

```
#### Generate Simulated data example ####
set.seed(30)
x=c(rnorm(50),rnorm(50,mean=5),rnorm(1,mean=15),rnorm(49,mean=5),rnorm(50,mean=4))
xcpt = cpt.mean(x,method='PELT') # Get the changepoints via PELT

#### Get the influence for both delete and outlier options ####
x.inf = influence(xcpt)

#### Parameter Stability plot ####
ParameterStability(x.inf, original.mean = rep(param.est(xcpt)$mean,
    times=diff(c(0,xcpt@cpts))), las = 1)
# note that the original mean is an n length vector and you can use the above code
# to get this from the original changepoint locations.
```

StabilityOverview

Stability Overview Graphic

Description

Plots the overview of the stability according to the influence measure.

StabilityOverview 11

Usage

```
StabilityOverview(data, original.cpts, influence,cpt.lwd=2,
  cpt.col=c("#009E73", "#E69F00", "#E41A1C"),cpt.lty=c("dashed","dotdash","dotted"),
  ylab=' ',xlab='Index', legend.args=list(display=TRUE,x="left",y=NULL,cex = 1,bty="n",
  horiz=TRUE,xpd=FALSE), ...)
```

Arguments

data	A vector containing the data on which you have run your changepoint method.
original.cpts	An ordered vector of the changepoint locations found by your favourite changepoint method.
influence	The influence as calculated the influence() function provided within this package. This is a list object.
cpt.lwd	The line width to be used when plotting the original.cpts on the data. Standard lwd values allowed. A single value is expected.
cpt.col	Colour of the original.cpts lines when plotted. We need three colours specified here, corresponding to the "stable", "unstable" and "outlier" categories respectively. Any values accepted by the col plotting argument are allowed.
cpt.lty	Line type of the original.cpts lines when plotted. We need three line types specified here, corresponding to the "stable", "unstable" and "outlier" categories respectively. Any values accepted by the 1ty plotting argument are allowed.
ylab, xlab	The labels for the x- and y-axis, character vector expected.
legend.args	These arguments are passed to the legend() function to control the legend position etc
	Any other arguments to be passed to the plot() function.

Details

This function creates a first summary graphic to display the results of a changepoint influence analysis. The graphic is a plot of the original data with the changepoints as vertical lines at their respective positions. The colour and line type of the changepoint vertical lines reflect their stability. The first value of their arguments denotes a stable changepoint - which appears at the same location in all influence segmentations. The second argument denotes an unstable changepoint - which doesn't appear at the same location in all influence segmentations, either it moves or is deleted. The third argument denotes changepoint locations which are deemed outliers as two changepoints occur at consecutive locations (surrounding the outlying observation).

Value

The function returns a plot and a list containing the labels of the original.cpts as either "stable", "unstable", or "outlier".

Author(s)

Rebecca Killick

12 welldata

References

Wilms I, Killick R, Matteson DS (2021) Graphical Influence Diagnostics for Changepoint Models, *ArxiV*: 2107.10572

See Also

```
influence, LocationStability,ParameterStability,InfluenceMap
```

Examples

```
#### Generate Simulated data example ####
set.seed(30)
x=c(rnorm(50),rnorm(50,mean=5),rnorm(1,mean=15),rnorm(49,mean=5),rnorm(50,mean=4))
xcpt = cpt.mean(x,method='PELT') # Get the changepoints via PELT

#### Get the influence for both delete and outlier options ####
x.inf = influence(xcpt)

#### Stability Dashboard ####
StabilityOverview(x,cpts(xcpt),x.inf,las=1,
legend.args=list(display=TRUE,x="topright",y=NULL,cex=1.5,bty="n",horiz=FALSE,xpd=FALSE))
```

welldata

Welllog data

Description

This data has been used in previous changepoint papers and is described and provided in "On-line inference for hidden Markov models via particle filters" by Fearnhead and Clifford in 2003. The data consists of measurements of the nuclear magnetic response of underground rocks.

Please note that this is the original data. The data analyzed in the majority of publications has been standardized and/or had the outliers removed. Papers typically only analyze a portion of the 4050 vector too.

Usage

welldata

Format

A vector of length 4050.

Source

https://doi.org/10.1111/1467-9868.00421

welldata 13

Examples

```
#### Load the data in the R package changepoint.influence ####
data("welldata")
welllog = welldata[1001:2000]
# Extract the mid section of the data as analyzed in other papers
n = length(welllog)
var = NULL; for (i in 30:1000){var[i]=var(welllog[(i-29):i])}
welllogs = welllog/sqrt(median(var, na.rm = TRUE))
# rescale the data to have unit variance across time,
# note that there may still be changes in variance across the series.
#### Apply PELT to the welllog data ####
out.PELT = cpt.mean(welllogs, method = 'PELT')
#### Calculate the influence measures ####
welllogs.inf = influence(out.PELT)
# the code extracts all the details of the original cpt.mean() function call
# and uses these in the calculation of the influence for the modified data.
#### Stability Dashboards ####
StabilityOverview(welllogs,cpts(out.PELT),welllogs.inf,las=1,ylab='Nuclear-Magnetic Response',
  legend.args=list(display=TRUE,x="bottomright",y=NULL,cex=1.5,bty="n",horiz=FALSE,xpd=FALSE))
# We can specify where the legend will sit in the graphic via the legend.args
# which are passed to the legend() function. We can also include additional
# arguments to pass to the plotting such as las=1 here.
#### Location Stability plot ####
exp.seg=LocationStability(cpts(out.PELT), welllogs.inf, type = 'Difference', cpt.lwd = 4, las = 1)
# Note that if the expected segmentation is not provided, it will be calcuated
# and then returned so that the user can avoid calculating this again in other plot calls.
#### Parameter Stability plot ####
ParameterStability(welllogs.inf, original.mean = rep(param.est(out.PELT)$mean,
  times=diff(c(0,out.PELT@cpts))), las = 1, ylab = 'Nuclear-Magnetic Response')
# Note that the original.mean argument is provided for each timepoint so is a length n vector.
#### Influence Map ####
welllogs.inf = influence(out.PELT, method = "delete")
inf.resid.del=InfluenceMap(cpts(out.PELT), welllogs.inf, data = welllogs, include.data = TRUE,
 ylab = 'Nuclear-Magnetic\n Response')
welllogs.inf = influence(out.PELT, method = "outlier")
inf.resid.out=InfluenceMap(cpts(out.PELT), welllogs.inf, data = welllogs, include.data = TRUE,
 ylab='Nuclear-Magnetic\n Response')
```

Index

* changepoint	changepoint.influence-package, 2
changepoint.influence-package, 2	
InfluenceMap, 4	influence, 6, 8, 10, 12
LocationStability, 7	InfluenceMap, 3, 4, 8, 10, 12
ParameterStability, 9	
StabilityOverview, 10	LocationStability, <i>3</i> , <i>6</i> , <i>7</i> , <i>10</i> , <i>12</i>
* datasets	ParameterStability, <i>3</i> , <i>6</i> , <i>8</i> , <i>9</i> , <i>12</i>
welldata, 12	Far affected Stability, $3, 0, 8, 9, 12$
* influence	StabilityOverview, 3, 6, 8, 10, 10
changepoint.influence-package, 2	
InfluenceMap, 4	welldata, 12
LocationStability, 7	welllog (welldata), 12
ParameterStability, 9	
StabilityOverview, 10	
* methods	
InfluenceMap, 4	
LocationStability, 7	
ParameterStability, 9	
StabilityOverview, 10	
* models	
InfluenceMap, 4	
LocationStability, 7	
ParameterStability, 9	
StabilityOverview, 10	
* package	
changepoint.influence-package, 2	
* plot	
InfluenceMap, 4	
LocationStability, 7	
ParameterStability, 9	
StabilityOverview, 10	
* segmentation	
InfluenceMap, 4	
LocationStability, 7	
ParameterStability, 9	
StabilityOverview, 10	
changepoint.influence	
<pre>(changepoint.influence-package),</pre>	
2	