Package 'akmeans'

February 19, 2019

Type Package

Title akmeans: 'Anchored' kmeans for Longitudinal Data

Version 0.1.0

Date 2019-02-06

Author Monsuru Adepeju [cre, aut], Samuel Langton [aut], Jon Bannister [aut]

Maintainer Monsuru Adepeju <monsuurg2010@gmail.com>

Description Advances an akmeans clustering technque and a stability-based quality criterion for longitudinal data. Also, contains functions for useful for the analysis of longitudinal data.

License GPL-2

Encoding UTF-8

LazyData TRUE

Imports devtools, Hmisc, ggplot2, rgdal, base, utils, reshape2, later, kml

Suggests knitr, rmarkdown

RoxygenNote 6.1.1

VignetteBuilder knitr

R topics documented:

akmeans_clust *akmeans_clust*

Description

GGGGGGGGG

Usage

```
akmeans_clust(dat = gm_crime_data, id_field = FALSE,
  init_centroids = "lpm", n_clusters = 3)
```

Arguments

dat	A matrix or data.frame with each row representing the trajectory of observations of a unique location. The columns show the observation at consecutive time steps.
id_field	Whether the first column is a unique (id) field. [default: FALSE]
init_centroids	initialisation method [default: "lpm" - linear partitioning medoids @seealso lpm_centroids]
n_clusters	number of clusters to generate [default (minimum value): 3]

Value

data_clusters_list

---

alphaLabel *Function to assign alphabetic labels to numeric cluster IDs*

Description

GGGGGGGGGGG

Usage

```
alphaLabel(x = clusters)
```

Arguments

x	A vector of numeric cluster ids

gm_crime_data *Sample crime dataset*

Description

Crime dataset of greater Manchester crime data aggregated at the LSOA geographical level data (Source: data.police.uk)

Usage

```
gm_crime_data
```

Format

A matrix

lpm_centroids *Linear Partition Medoids (LPM) Centroids*

Description

Linear Partition Medoids (LPM) Centroids

Usage

```
lpm_centroids(dat, id_field2 = FALSE, n_centroids = 3)
```

Arguments

dat	A matrix or data.frame with each row representing the trajectory of observations of a unique location. The columns show the observation at consecutive time steps.
id_field2	Whether the first column is a unique (id) field. [default: FALSE]
n_centroids	Number of initial (linear) centroids to generate based on lpm technique (Adepeju et al. 2019, under review)

Value

l_centroids

null_filler *Data Imputing*

Description

This function replaces any cells with the entry 'NA' or 'Inf' in a matrix or data.frame with the 'Mean', 'Minimum' or 'Maximum' value of either the column or row in which the cell is located.

Usage

```
null_filler(dat, id_field = FALSE, replace_with = "Mean_col")
```

Arguments

dat	A matrix or data.frame with each row representing the trajectory of a unique location. The columns show the observation at consecutive time steps.
id_field	Whether the first column is a unique (id) field. [default: FALSE]
replace_with	Values to replace with [Values: "Mean_col", "Min_col", "Max_col, "Mean_row", "Min_row" or "Max_row"]. The default is "Mean_col", meaning to imput the average values of the field in which the cell is located.

Value

datF

outlierDetect *Outlier detection in longitudinal or repeated observations*

Description

Detect outlier in a longitudinal or repeated data. This function identify the outlier observations according to a specified method. A matrix, 'outlier_mat', is created with entries 'TRUE' or 'FALSE' indicating whether or not an observation is an outlier. The final list of outlier trajectories is determined by the 'hortz_tolerance' parameter i.e. how many observation in a trajectory exceed the 'threshold' value.

Usage

```
outlierDetect(dat, id_field = FALSE, method = "quantile",
  threshold = 0.95, hortz_tolerance = 1, replace_with = "Mean_row")
```

Arguments

dat	A matrix or data.frame with each row representing the trajectory of observations of a unique location. The columns show the observation at consecutive time steps.
id_field	Whether the first column is a unique (id) field. [default: FALSE]
method	Specify the method for identifying the outlier. Available methods: (1) "quantile" (2) "manual" - a user-defined value

threshold — Value in which an observation must exceed in order to be flagged as outlier. Depending on the method specified: (1) for "quantile" method, enter a numeric vector of probabilities with values in [0,1], (2) for "Manual" method: a user-specified value.

hortz_tolerance — Specifying the number of observations of a trajectory that have to exceed the cut-off 'threshold' value in order for the trajectory to be flagged as outlier. [default: 1]

replace_with — Value to replace the outlier observation with. Values to replace with [Values: "Mean_col" or "Mean_row"]. The default is "Mean_row", meaning to imput the average values of the field in which the observation is located.

Value

dat_

---

plot_clust *To plot the clusters*

Description

To plot the clusters

Usage

```
plot_clust(data_clusters_list, id_field = TRUE)
```

Arguments

data_clusters_list — A data.frame of clusters from akmeans_clust, in which the last column represents alphabetical cluster ids (labels)

id_field — Whether the first column is a unique (id) field. [default: TRUE]

Value

data_clusters_list

---

props *Function to convert counts or rates to proportion*

Description

Function to convert counts or rates to proportion

Usage

```
props(rates, id_field = FALSE)
```

Arguments

rates	A matrix or data.frame with each row representing the trajectory of observations of a unique location. The columns show the observation at consecutive time steps.
id_field	Whether the first column is a unique (id) field. [default: FALSE]

Value

props

---

qpm_centroids *Quadratic Partition Medoids (QPM) Centroids*

Description

Quadratic Partition Medoids (QPM) Centroids

Usage

```
qpm_centroids(dat, n_centroids = 3, id_field = FALSE)
```

Arguments

dat	A matrix or data.frame with each row representing the trajectory of observations of a unique location. The columns show the observation at consecutive time steps.
n_centroids	Number of initial (quadratic) centroids to generate based on the qpm method (See attached Vignette)
id_field	Whether the first column is a unique (id) field. [default: FALSE]

Value

q_centroids

---

whiteSpaces *Function to remove whitespaces in data entries*

Description

Function to remove whitespaces in data entries

Usage

```
whiteSpaces(dat, head = TRUE)
```

Arguments

dat	A matrix or data.frame
head	If column names exist

Value

dat_Cleaned

Index