

Package ‘CLARA.seq’

April 26, 2021

Type Package

Title CLARA Clustering

Version 0.1.0

Author Louis Tochon (with Matthias Studer)

Maintainer Louis Tochon <louis.tochon@etu.unige.ch>

Description

Clustering algorithm (CLARA) to handle big datasets and index to evaluate the quality of the partition. This package works with the help of TraMineR's package for distance calculations.

Imports TraMineR, cluster, dplyr, doParallel

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

R topics documented:

clara_clust	1
db_index	3

Index	4
--------------	----------

clara_clust	<i>CLARA clustering</i>
-------------	-------------------------

Description

With the help of TraMineR package, CLARA clustering provide a clustering of big dataset. The main objective is to cluster state sequences with the "LCS" distance calculation method to find the best partition in N clusters.

Usage

```
clara_clust(
  data,
  nb_sample = 100,
  size_sample = 40 + 2 * nb_cluster,
  nb_cluster = 4,
  method = "LCS",
  plot = FALSE,
  improve_method = "Random",
  nbessai_improved = 20,
  percent_improve = 0.1
)
```

Arguments

<code>data</code>	The dataset to use. In case of sequences, use <code>seqdef</code> (from <code>TraMineR</code> package) to create such an object.
<code>nb_sample</code>	The number of subsets to test.
<code>size_sample</code>	The size of each subset
<code>nb_cluster</code>	The number of medoids
<code>method</code>	The calculation method to compute the distance matrix. (See the function <code>seqdist</code> in <code>TraMineR</code> package)
<code>plot</code>	Boolean variable to plot the result of clustering
<code>improve_method</code>	Addind an improve method to find better subsets to test (increase the calculation cost)
<code>nbessai_improved</code>	The number of tries to find the best subset to test
<code>percent_improve</code>	The percent of <code>nb_sample</code> using <code>improve_method</code>

Value

An object with the data, the medoids id, the clustering and the distance matrix

Examples

```
data <- data(biofam)
#Basic CLARA Computing
my_cluster <- clara_clust(data, method="LCS")

#Improved CLARA Computing
my_cluster <- clara_clust(data, method="LCS", improve_method="Silhouette", nbessai_improved=30)
```

db_index	<i>Davies Bouldin Index</i>
----------	-----------------------------

Description

Implementation of Davies-Bouldin index to evaluate the quality of CLARA Clustering.

Usage

```
db_index(seq_obj, method, diss = TRUE, plot = TRUE)
```

Arguments

seq_obj	The object generated with CLARA Clustering
method	The calculation method to compute the distance matrix. (See the function seqdist in TraMineR package)
diss	Boolean to express if the parameter diss from CLARA.seq clustering has been returned
plot	Boolean variable to plot the result

Value

The value of the index

Examples

```
my_index <- db_index(my_cluster, "LCS", diss = FALSE)
```

Index

clara_clust, [1](#)

db_index, [3](#)