Package 'ODMeans'

March 18, 2024

Type Package

Title OD-Means: k-Means for Origin-Destination

Version 0.2.1

Maintainer Sebastian Moreno <sebastian.moreno.araya@gmail.com>

Description OD-means is a hierarchical adaptive k-means algorithm based on origin-destination pairs. In the first layer of the hierarchy, the clusters are separated automatically based on the variation of the within-cluster distance of each cluster until convergence. The second layer of the hierarchy corresponds to the sub clustering process of small clusters based on the distance between the origin and destination of each cluster.

License GPL (>= 3)

Encoding UTF-8

LazyData false

RoxygenNote 7.3.1

Imports geosphere, ggplot2, stats, ggmap, ggrepel

Collate 'ODMeansSampleData.R' 'ODMeansTaxiData.R' 'first_hierarchy.R' 'second_hierarchy.R' 'odmeans.R' 'odmeans_graph.R'

Depends R (>= 3.5.0)

NeedsCompilation no

Author Sebastian Moreno [aut, cre],
Cristobal Heredia [aut],
Wilfredo Yushimito [ths],
Gonzalo Huerta [ths]

Repository CRAN

Date/Publication 2024-03-18 22:20:07 UTC

R topics documented:

`first_hierarchy` *First Hierarchy Function*

Description

First Hierarchy Function

Usage

```
first_hierarchy(data, numK, limitsSeparation, maxDist, kmeans_pp = FALSE)
```

Arguments

`data`	A data frame with four columns: Initial Latitude | Initial Longitude | Final Latitude | Final Longitude
`numK`	Initial number of clusters in the first call of k-means in the global hierarchy.
`limitsSeparation`	Within cluster distance threshold to determine if a global cluster must be separated into two new clusters.
`maxDist`	Meter distance threshold used to re-estimate centroids in global hierarchy.
`kmeans_pp`	Boolean value, if TRUE it initialize centroids using kmeans++.

Value

Returns an S3 class object similar to kmeans S3 Class, with eight properties.

Examples

```
data(ODMeansTaxiData)
first_hierarchy_data = first_hierarchy(ODMeansTaxiData, 10, 300, 1000)
```

odmeans *ODMeans Function*

Description

ODMeans Function

Usage

```
odmeans(
  data,
  numKGlobal,
  limitSeparationGlobal,
  maxDistGlobal,
  distHierarchical,
  numKLocal,
  limitSeparationLocal,
  maxDistLocal,
  kmeans_pp = FALSE
)
```

Arguments

`data`	A data frame with four columns: Initial Latitude | Initial Longitude | Final Latitude | Final Longitude
`numKGlobal`	Initial number of clusters in the first call of k-means in the global hierarchy.
`limitSeparationGlobal`	Within cluster distance threshold to determine if a global cluster must be separated into two new clusters.
`maxDistGlobal`	Meter distance threshold used to re-estimate centroids in global hierarchy.
`distHierarchical`	Meter distance threshold between origin and destination to generate new local clusters from a first layer cluster
`numKLocal`	Initial number of clusters in the first call of k-means in the local hierarchy.
`limitSeparationLocal`	Within cluster distance threshold to determine if a local cluster must be separated into two new clusters.
`maxDistLocal`	Meter distance threshold used to re-estimate centroids in local hierarchy.
`kmeans_pp`	Boolean value, if TRUE it initialize centroids using kmeans++.

Value

Returns an S3 class object similar to kmeans S3 Class, with eight properties.

Examples

```
data(ODMeansTaxiData)
odmeans_data = odmeans(ODMeansTaxiData, 10, 300, 1000, 2200, 3, 50, 100)
```

`ODMeansSampleData` *Origin-Destination points*

Description

A synthetic data frame containing 1700 Origin-Destination points.

Usage

```
data(ODMeansSampleData)
```

Format

A data frame with 10000 rows and 5 variables:

OriginLatitude Consists of the origin latitude dimension

OriginLongitude Consists of the origin longitude dimension

DestinationLatitude Consists of the destination latitude dimension

DestinationLongitude Consists of the destination longitude dimension

original_cluster Original cluster of the points when it was created ...

Source

Synthetic data

`ODMeansTaxiData` *Origin-Destination Taxi data*

Description

The data frame contains the 452,166 trips collected for the months of March (2014 to 2016), July (2014 to 2016), and November (2014 and 2015). The data points are the taxi's initial and ending location based on latitude and longitude.

Usage

```
data(ODMeansTaxiData)
```

Format

A data frame with 452,166 rows and 4 variables:

OriginLatitude Origin latitude dimension

OriginLongitude Origin longitude dimension

DestinationLatitude Destination latitude dimension

DestinationLongitude Destination longitude dimension ...

Source

Fantaxico, Fermanti Servicios de Ingeniería S.A, Santiago, Chile.

---

odmeans_graph *Graph ODMeans Function*

Description

Graph ODMeans Function

Usage

```
odmeans_graph(
  odmeans_data,
  title = "ODMeans Graph",
  maptype = "roadmap",
  zoom = 4,
  add_cluster = TRUE
)
```

Arguments

odmeans_data	It receives an object from S3 ODMeans class. However, it can also work with objects from similar classes like S3 k-Means
title	It receives an string, and corresponds to the title of the plot.
maptype	It receives a string with the type of the map. Depending on the map selected, it will change the background of it. The possible values are: "terrain", "satellite", "roadmap", "hybrid".
zoom	An integer from 3 (continent) to 21 (building), which controls the level of zoom applied to the map.
add_cluster	Receives TRUE or FALSE value. When True, it adds the number of the cluster to the arrows.

Value

A ggplot graph showing a map with the centers of the clusters.

Examples

```
data(ODMeansTaxiData)
odmeans_data = odmeans(ODMeansTaxiData, 10, 300, 1000, 2200, 3, 50, 100)
odmeans_plot = odmeans_graph(odmeans_data, "ODMeans Taxi Graph", "roadmap", 11, FALSE)
```

`second_hierarchy` *Second Hierarchy Clusters*

Description

Second Hierarchy Clusters

Usage

```
second_hierarchy(
  data,
  Kcluster,
  distHierarchical,
  numKLocal,
  limitSeparationLocal,
  maxDistLocal
)
```

Arguments

`data`	A data frame with four columns: Initial Latitude | Initial Longitude | Final Latitude | Final Longitude
`Kcluster`	An ODMeans structure, result of function first_hierarchy.
`distHierarchical`	Meter distance threshold between origin and destination to generate new local clusters from a first layer cluster.
`numKLocal`	Initial number of clusters in the first call of k-means in the local hierarchy.
`limitSeparationLocal`	Within cluster distance threshold to determine if a local cluster must be separated into two new clusters.
`maxDistLocal`	Meter distance threshold used to re-estimate centroids in local hierarchy.

Value

Returns an S3 class object similar to kmeans S3 Class, with eight properties.

Examples

```
data(ODMeansTaxiData)
first_hierarchy_data = first_hierarchy(ODMeansTaxiData, 10, 500, 1000)
second_hierarchy_data = second_hierarchy(ODMeansTaxiData, first_hierarchy_data, 2200, 3, 50, 100)
```

Index