Package ‘RadialVisGadgets’

Type Package
Title Interactive Gadgets for Radial Visualization Approaches
Version 0.2.0
URL https://github.com/jmatute/RadialShinyGadgets
Imports import, ggplot2, tidyr, dplyr, miniUI, shiny, shinyjs, caret, rlang, shinyscreenshot
Maintainer José Matute <jmatuteflores@gmail.com>
Description Shiny-based interactive gadgets of radial visualization methods and extensions thereof.
License MIT + file LICENSE
Encoding UTF-8
LazyData true
RoxygenNote 7.1.1
Suggests knitr, rmarkdown, datasets, cIValid
VignetteBuilder knitr
NeedsCompilation no
Author José Matute [aut, cre]
Repository CRAN
Date/Publication 2020-12-11 10:20:08 UTC

R topics documented:

RadViz .......................................................... 2
StarCoordinates ............................................. 3

Index 6
Description

Creates a RShiny Gadget for RadViz

Usage

RadViz(df, color = NULL)

Arguments

df A dataframe with the data to explore. It should contain only numeric columns (with the exception of the label column).
color column where labels from the data are extracted.

Details

RadViz’s goal is to generate a configuration which reveals the underlying nature of the data for cluster analysis, outlier detection, and exploratory data analysis, e.g., by investigating the effect of specific dimensions on the separation of the data. Each dimension is assigned to a point known as dimensional anchors across a unit-circle. Each sample is projected according to the relative attraction to each of the anchors.

It is defined for multidimensional numerical data sets \( X = \{ p_1, \ldots, p_N \} \), for \( N \) data points \( x_i \in \mathbb{R}^d \) of dimensionality \( d \). Let \( A = \{ a_1, \ldots, a_d \} \), be a set of (typically 2D) anchors, each corresponding to one of the \( d \) dimensions. The projection \( p'_i \in \mathbb{R}^2 \), of a multidimensional point \( p_i = (p_{i1}, \ldots, p_{id}) \in \mathbb{R}^d \), in SC is then defined as:

\[
x'_i = \frac{\sum_{j=1}^{d} a_j g_j(p_i)}{\sum_{j=1}^{d} a_j},
\]

with

\[
g_j(p_i) = \frac{p_{ij} - \min_j}{\max_j - \min_j},
\]

and \((\min_j, \max_j)\), denoting the value range of dimension \( j \). The dimensional anchors can be moved either interactively or algorithmically to reveal different meaningful patterns in the dataset.

Value

A list location of the anchors, coordinates of the projected samples and a logical vector with the selected samples

References

Examples

```r
if (interactive()) {
  library(RadialVisGadgets)
  library(datasets)
  data(iris)
  RadViz(iris, "Species")
}
```

---

**StarCoordinates**  
*Star Coordinates Gadget*

---

**Description**

Creates a RShiny Gadget for Star Coordinates

**Usage**

```r
StarCoordinates(
  df, 
  color = NULL, 
  approach = "Standard", 
  numericRepresentation = TRUE, 
  meanCentered = TRUE, 
  projMatrix = NULL, 
  clusterFunc = NULL
)
```

**Arguments**

df: A dataframe with the data to explore. It should contain only numeric or factor columns.
dcolor: column where labels from the data are extracted.
approach: Standard approach as defined by Kandogan, or Orthographic Star Coordinates (OSC) with a recondition as defined by Lehmann and Thiesel
numericRepresentation: if true attempt to convert all factors to numeric representation, otherwise used mixed representation as defined in Hinted Star Coordinates
meanCentered: center the projection at the mean of the values. May allow for easier value estimation
projMatrix: a pre-defined projection matrix as an initial configuration. Should be defined in the same fashion as the output
clusterFunc: function to define hints, assume increase in value of the function is an increase in quality of the projection. The function will be called with two parameters (points, labels)
Details

Star Coordinates’ (SC) goal is to generate a configuration which reveals the underlying nature of the data for cluster analysis, outlier detection, and exploratory data analysis, e.g., by investigating the effect of specific dimensions on the separation of the data. Traditional SC are defined for multi-dimensional numerical data sets \( X = \{ p_1, \ldots, p_N \} \), for \( N \) data points \( x_i \in \mathbb{R}^d \) of dimensionality \( d \). Let \( A = \{ a_1, \ldots, a_d \} \), be a set of (typically 2D) vectors, each corresponding to one of the \( d \) dimensions. The projection \( p'_i \in \mathbb{R}^2 \), of a multidimensional point \( p_i = (p_{i1}, \ldots, p_{id}) \in \mathbb{R}^d \), in SC is then defined as:

\[
x'_i = \sum_{j=1}^{d} a_j g_j(p_i),
\]

with

\[
g_j(p_i) = \frac{p_{ij} - \min_j}{\max_j - \min_j},
\]

and \((\min_j, \max_j)\), denoting the value range of dimension \( j \).

In the case of categorical dimensions, the values when numericRepresentation= TRUE are mapped into numerical type i.e. as.numeric() However equally spaced categorical points may not reflect the true nature of the data. Instead, a frequency-based representation may be applied for individual data points. Assuming a categorical dimension \( j \), we calculate the frequency \( f_{jk} \), of each category \( k \) of dimension \( j \). The respective axis vector \( a_j \), is then divided into according blocks, whose size represent the relative frequency (or probability) \( \sum_{l=1}^{m} f_{jl} \), of each of the \( m \) categories of dimension \( j \).

In summary, given an order for each categorical dimension, the Equation \( g() \), above can be extended to SC for mixed data by:

\[
g_j(x_i) = F_j(x_{ij}) - \frac{P_j(x_{ij})}{2},
\]

if categorical/ordinal

\[
g_j(x_i) = \frac{x_{ij} - \min_j}{\max_j - \min_j},
\]

if numerical

where \( F_j \), is the cumulative density function for (categorical/ordinal) dimension \( j \) and \( P_j \), its probability function.

Value

A list with the projection matrix, coordinates of the projected samples and a logical vector with the selected samples

References


Examples

```r
if (interactive()) {
  library(RadialVisGadgets)
  library(datasets)
  data(iris)
  StarCoordinates(iris, "Species")
}
```
Index

RadViz, 2

StarCoordinates, 3