Package 'mldr'

February 13, 2015

Title Exploratory Data Analysis and Manipulation of Multi-Label Data Sets

Version 0.1.63

Date 2015-02-08

Description Exploratory data analysis and manipulation functions for multi-label data sets along with interactive Shiny application to ease their use.

Depends R (>= 3.0.0),
shiny (>= 0.11),
XML,
circlize

License GPL (>= 3) | file LICENSE

LazyData true

R topics documented:

+.mldr — *Generates a new mldr object joining the rows in the two mldrs given as input*

Description

Generates a new mldr object joining the rows in the two mldrs given as input

Usage

```
## S3 method for class mldr
mldr1 + mldr2
```

Arguments

mldr1	First mldr object to join
mldr2	Second mldr object to join

Value

a new mldr object with all rows in the two parameters

==.mldr — *Checks if two mldr objects have the same structure*

Description

Checks if two mldr objects have the same structure

Usage

```
## S3 method for class mldr
mldr1 == mldr2
```

Arguments

mldr1	First mldr object to compare
mldr2	Second mldr object to compare

Value

TRUE if the two mldr objects have the same structure, FALSE otherwise

birds *birds*

Description

birds dataset.

Usage

```
birds
```

Format

An mldr object with 645 instances, 279 attributes and 19 labels

Source

F. Briggs, Yonghong Huang, R. Raich, K. Eftaxias, Zhong Lei, W. Cukierski, S. Hadley, A. Hadley, M. Betts, X. Fern, J. Irvine, L. Neal, A. Thomas, G. Fodor, G. Tsoumakas, Hong Wei Ng, Thi Ngoc Tho Nguyen, H. Huttunen, P. Ruusuvuori, T. Manninen, A. Diment, T. Virtanen, J. Marzat, J. Defretin, D. Callender, C. Hurlburt, K. Larrey, M. Milakov. "The 9th annual MLSP competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment", in proc. 2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)

Examples

```
summary(birds)
birds$labels
```

emotions *emotions*

Description

emotions dataset.

Usage

```
emotions
```

Format

An mldr object with 593 instances, 78 attributes and 6 labels

Source

K. Trohidis, G. Tsoumakas, G. Kalliris, I. Vlahavas. "Multilabel Classification of Music into Emotions". Proc. 2008 International Conference on Music Information Retrieval (ISMIR 2008), pp. 325-330, Philadelphia, PA, USA, 2008

Examples

```
summary(emotions)
emotions$labels
```

`genbase` *genbase*

Description

genbase dataset.

Usage

```
genbase
```

Format

An mldr object with 662 instances, 1213 attributes and 27 labels

Source

S. Diplaris, G. Tsoumakas, P. Mitkas and I. Vlahavas. Protein Classification with Multiple Algorithms, Proc. 10th Panhellenic Conference on Informatics (PCI 2005), pp. 448-456, Volos, Greece, November 2005

Examples

```
summary(genbase)
genbase$labels
```

`mldr` *Creates an object representing a multilabel dataset*

Description

Reads a multilabel dataset from a file and returns an `mldr` object containing the data and additional measures. The file has to be in ARFF format. The label information could be in a separate XML file (MULAN style) or in the the arff header (MEKA style)

Usage

```
mldr(filename = NULL, use_xml = TRUE, auto_extension = TRUE,
  xml_file = NULL)
```

Arguments

filename	Name of the dataset
use_xml	Specifies whether to use an associated XML file to identify the labels. Defaults to TRUE
auto_extension	Specifies whether to add the '.arff' and '.xml' extensions to the filename where appropriate. Defaults to TRUE
xml_file	Path to the XML file. If not provided, the filename ending in ".xml" will be assumed

Value

An mldr object containing the multilabel dataset

See Also

mldr_from_dataframe, summary.mldr

Examples

```
library(mldr)
## Not run:
# Read "yeast.arff" and labels from "yeast.xml"
mymld <- mldr("yeast")

# Read "yeast-tra.arff" and labels from "yeast.xml"
mymld <- mldr("yeast-tra", xml_file = "yeast.xml")

# Read MEKA style dataset, without XML file and giving extension
mymld <- mldr("IMDB.arff", use_xml = FALSE, auto_extension = FALSE)

## End(Not run)
```

---

mldrGUI *Launches the web-based GUI for mldr*

Description

Loads an interactive user interface in the web browser, built using R shiny.

Usage

```
mldrGUI()
```

Details

The **mldr** package provides a basic, Shiny-based GUI to work with multilabel datasets. You have to install the **shiny** package to be able to use this GUI.

The user interface allows working with any of the previous loaded datasets, as well as loading new ones. The GUI is structured into the following pages:

- **Main:** This page is divided into two sections. The one at the left can be used to choose a previously loaded dataset, as well as to load datasets from files. The right part shows some basic statistics about the selected multilabel dataset.
- **Labels:** This page shows a table containing for each label its name, index, count, relative frequency and imbalance ratio (IRLbl). The page also includes a bar plot of the label frequency. The range of labels in the plot can be customized.
- **Labelsets:** This page shows a table containing for each labelset its representation and a counter.
- **Attributes:** This page shows a table containing for each attribute its name, type and a summary of its values.
- **Concurrence:** This page shows for each label the number of instances in which it appears and its mean SCUMBLE measure, along with a plot that shows the level of concurrence among the selected labels. Clicking the labels in the table makes it possible to add/remove them from the plot.

The tables shown in these pages can be sorted by any of its fields, as well as filtered. The content of the tables can be copied to clipboard, printed and saved in CSV and Microsoft Excel format.

Value

Nothing

Examples

```
## Not run:
library(mldr)
mldrGUI()

## End(Not run)
```

---

mldr_from_dataframe *Generates an mldr object from a data.frame and a vector with label indices*

Description

This function creates a new `mldr` object from the data stored in a `data.frame`, taking as labels the columns pointed by the indexes given in a vector.

Usage

```
mldr_from_dataframe(dataframe, labelIndices, name = NULL)
```

Arguments

`dataframe`	The `data.frame` containing the dataset attributes and labels.
`labelIndices`	Vector containing the indices of attributes acting as labels. Usually the labels will be at the end (right-most columns) or the beginning (left-most columns) of the `data.frame`
`name`	Name of the dataset. The name of the dataset given as first parameter will be used by default

Value

An mldr object containing the multilabel dataset

See Also

mldr, summary.mldr

Examples

```
library(mldr)

df <- data.frame(matrix(rnorm(1000), ncol = 10))
df$Label1 <- c(sample(c(0,1), 100, replace = TRUE))
df$Label2 <- c(sample(c(0,1), 100, replace = TRUE))
mymldr <- mldr_from_dataframe(df, labelIndices = c(11, 12), name = "testMLDR")

summary(mymldr)
```

---

mldr_transform *Transformns an MLDR into binary or multiclass datasets*

Description

Transforms an mldr object into one or serveral binary or multiclass datasets, returning them as data.frame objects

Usage

```
mldr_transform(mldr, type = "BR", labels)
```

Arguments

mldr	The mldr object to transform
type	Indicates the type of transformation to apply. Possible types are: • "BR" Produces one or more binary datasets, each one with one label • "LP" Produces a multiclass dataset using each labelset as class label
labels	Vector with the label indexes to include in the transformation. All labels will be used if not specified

Value

A list of data.frames containing the resulting datasets (for BR) or a data.frame with the dataset (for LP). The result is no longer an mldr object, but a plain data.frame

Examples

```
library(mldr)
emotionsbr <- mldr_transform(emotions, type = "BR")
emotionslp <- mldr_transform(emotions, type = "LP")
```

plot.mldr *Generates graphic representations of an mldr object*

Description

Generates graphic representations of an `mldr` object

Usage

```
## S3 method for class mldr
plot(x, type = "LC", labelCount, labelIndices, title = NULL,
  ...)
```

Arguments

`x`	The mldr object whose features are to be drawn
`type`	Indicates the type of plot to be produced. Possible types are: • `"LC"` Draws a circular plot with sectors representing each label and links between them depicting label co-occurrences • `"LH"` for label histogram • `"LB"` for label bar plot • `"CH"` for cardinality histogram • `"AT"` for attributes by type pie chart • `"LSH"` for labelset histogram • `"LSB"` for labelset bar plot
`labelCount`	Samples the labels in the dataset to show information of only `labelCount` of them
`labelIndices`	Establishes the labels to be shown in the plot
`title`	A title to be shown above the plot. Defaults to the name of the dataset passed as first argument
`...`	Additional parameters to be given to barplot, hist, etc.

Examples

```
library(mldr)
## Not run:
# Label concurrence plot
plot(genbase, type = "LC") # Plots all labels
plot(genbase) # Same as above
plot(genbase, title = "genbase dataset") # Changes the title
plot(genbase, labelCount = 10) # Randomly selects 10 labels to plot
plot(genbase, labelIndices = genbase$label$index[1:10]) # Plots info of first 10 labels

# Label histogram plot
plot(emotions, type = "LH")

# Cardinality histogram plot
plot(emotions, type = "CH")

# Attributes by type
```

```
plot(emotions, type = "AT", cex = 0.85)

# Labelset histogram
plot(emotions, type = "LSH")

## End(Not run)
```

`print.mldr` *Prints the mldr content*

Description

Prints the `mldr` object data, including input attributs and output labels

Usage

```
## S3 method for class mldr
print(x, ...)
```

Arguments

`x`	Object whose data are to be printed
`...`	Additional parameters to be given to print

See Also

`mldr`, `summary.mldr`

Examples

```
library(mldr)

emotions
print(emotions) # Same as above
```

`summary.mldr` *Provides a summary of measures about the mldr*

Description

Prints a summary of the measures obtained from the `mldr` object

Usage

```
## S3 method for class mldr
summary(object, ...)
```

Arguments

`object`	Object whose measures are to be printed
`...`	Additional parameters to be given to print

See Also

mldr, print.mldr

Examples

```
library(mldr)

summary(emotions)
```

---

write_arff *Writes an* mldr *object to a file*

Description

Save the mldr content to an ARFF file and the label data to an XML file

Usage

```
write_arff(obj, filename, write.xml = FALSE)
```

Arguments

obj	The mldr object whose content is going to be written
filename	Base name for the files (without extension)
write.xml	TRUE or FALSE, stating if the XML file has to be written

See Also

mldr_from_dataframe, mldr

Examples

```
library(mldr)

write_arff(emotions, "myemotions")
```

[.mldr *Filter rows in a*mldr *returning a new* mldr

Description

Generates a new mldr object containing the selected rows from an existent mldr

Usage

```
## S3 method for class mldr
mldrObject[rowFilter = T]
```

Arguments

mldrObject	Original mldr object from which some rows are going to be selected
rowFilter	Expression to filter the rows

Value

A new mldr object with the selected rows

See Also

mldr_from_dataframe, ==.mldr, +.mldr

Examples

```
library(mldr)

highlycoupled <- genbase[.SCUMBLE > 0.05] # Select instances with highly imbalanced coupled labels
summary(highlycoupled)   # Compare the selected instances
summary(genbase)         # with the traits of the original MLD
```

Index