

Package ‘rsubgroup’

February 20, 2015

Type Package

Title Subgroup Discovery and Analytics

Version 0.6

Date 2014-09-10

Author Martin Atzmueller

Maintainer Martin Atzmueller <martin@atzmueller.net>

Description A collection of efficient and effective tools and algorithms for subgroup discovery and analytics. The package integrates an R interface to the org.vikamine.kernel library of the VIKAMINE system (<http://www.vikamine.org>) implementing subgroup discovery, pattern mining and analytics in Java.

Classification/ACM G.4, H.2.8, I.5.1

License GPL (>= 3)

Depends R (>= 2.10), methods, rJava (>= 0.6.3), foreign (>= 0.8.40)

SystemRequirements Java (>= 6.0)

Collate 'AAAonLoad.R' 'randomSeed.R' 'classes.R' 'subgroup.R'

URL <http://www.rsubgroup.org>

Repository CRAN

Repository/R-Forge/Project subgroup

Repository/R-Forge/Revision 51

Repository/R-Forge/DateTimeStamp 2014-09-10 15:40:17

Date/Publication 2014-09-11 11:00:24

NeedsCompilation no

R topics documented:

as.target	2
CreateSDTask	3
credit.data	3

DiscoverSubgroups	4
DiscoverSubgroupsByTask	5
Pattern-class	5
rsubgroup	6
SDTaskConfig-class	7
ToDataFrame	8
Index	9

as.target	<i>Constructs a target variable (for subgroup discovery)</i>
-----------	--

Description

Constructs a target variable, i.e., an object suitable to be passed to DiscoverSubgroups or CreateSDTask.

Usage

```
as.target(attribute, value=NULL)
```

Arguments

attribute	The attribute of the target variable.
value	For binary targets, the respective attribute value; the value is NULL for numeric targets.

See Also

[DiscoverSubgroups](#).

Examples

```
# creating a target variable
# binary:
as.target("class", "true")

#numeric:
as.target("numeric_class")
```

CreateSDTask	<i>Creates a Subgroup Discovery Task</i>
--------------	--

Description

Performs subgroup discovery according to the given task.

Usage

```
CreateSDTask(source, target, config=new("SDTaskConfig"))
```

Arguments

source	a data.frame or the a character string giving the filename of an ARFF file to use.
target	the target variable (constructed by <code>as.target</code>) to consider for subgroup discovery.
config	an instance of <code>SDTaskConfig</code> providing various parameters for subgroup discovery.

See Also

[DiscoverSubgroups](#). [DiscoverSubgroupsByTask](#) [SDTaskConfig](#)

Examples

```
# creating a task
data(credit.data)

# task with binary target
task <- CreateSDTask(credit.data, as.target("class", "good"))

# task with numeric target
taskNum <- CreateSDTask(credit.data, as.target("credit_amount"))
```

credit.data	<i>Statlog (German Credit Data) Data Set</i>
-------------	--

Description

This dataset classifies people described by a set of attributes as good or bad credit risks.

Usage

```
data(credit.data)
```

Format

A vector containing 1000 observations.

Source

UCI Repository, [https://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)).

DiscoverSubgroups *Performs Subgroup Discovery*

Description

Performs subgroup discovery according to the given target and the configuration on the data.

Usage

```
DiscoverSubgroups(source, target, config=new("SDTaskConfig"), as.df=FALSE)
```

Arguments

source	a data.frame or the a character string giving the filename of an ARFF file to use.
target	the target variable (constructed by as.target) to consider for subgroup discovery.
config	an instance of SDTaskConfig providing various parameters for subgroup discovery.
as.df	TRUE, if the result patterns should be returned as a data.frame using ToDataFrame

See Also

[DiscoverSubgroupsByTask](#). [as.target](#) [CreateSDTask](#) [SDTaskConfig](#)

Examples

```
# subgroup discovery on a data.frame, for binary target
data(credit.data)
result1 <- DiscoverSubgroups(
  credit.data, as.target("class", "good"), new("SDTaskConfig",
  attributes=c("checking_status", "credit_amount", "employment", "purpose")))
result2 <- DiscoverSubgroups(
  credit.data, as.target("class", "good"), new("SDTaskConfig",
  attributes=c("checking_status", "employment")))

ToDataFrame(result1)
ToDataFrame(result2)

# subgroup discovery for numeric target variable
result3 <- DiscoverSubgroups(
  credit.data, as.target("credit_amount"), new("SDTaskConfig",
  attributes=c("checking_status", "employment")))

ToDataFrame(result3)
```

`DiscoverSubgroupsByTask`*Performs Subgroup Discovery for a given Task*

Description

Performs subgroup discovery according to the given task.

Usage

```
DiscoverSubgroupsByTask(task, as.df=FALSE)
```

Arguments

`task` a subgroup discovery task constructed by `CreateSDTask`.
`as.df` TRUE, if the result patterns should be returned as a data.frame using [ToDataFrame](#)

See Also

[DiscoverSubgroups.CreateSDTask](#)

Examples

```
# creating a task
data(credit.data)
task <- CreateSDTask(
  credit.data, as.target("class", "bad"), new("SDTaskConfig",
  attributes=c("checking_status", "employment")))
taskNum <- CreateSDTask(
  credit.data, as.target("credit_amount"), new("SDTaskConfig",
  attributes=c("checking_status", "employment")))

# running the tasks
DiscoverSubgroupsByTask(task)
DiscoverSubgroupsByTask(taskNum)
```

`Pattern-class`*Class "Pattern" — A Simple Subgroup Description Container*

Description

A Simple Container holding the results (subgroups, description and parameters) for the Subgroup and Pattern Mining Algorithms

Objects from the Class

Objects are created by calls of the form `new("Pattern", ...)`.

Slots

description: The subgroup description, as a character vector.

quality: The numeric value denoting the quality of the subgroup pattern as determined by the applied quality function.

size: The size of the subgroup.

parameters Additional quality parameters of the subgroup.

See Also

[DiscoverSubgroups](#). [DiscoverSubgroupsByTask](#) [CreateSDTask](#)

rsubgroup

rsubgroup Package - Algorithms and Tools for Efficient Subgroup Discovery and Analytics

Description

The rsubgroup package contains a set of efficient and effective tools and algorithms for subgroup discovery and analytics.

Details

Package:	rsubgroup
Type:	Package
Version:	0.6
Date:	2014-09-10
License:	GPL (>= 3)
LazyLoad:	yes

Author(s)

Martin Atzmueller

Maintainer: Martin Atzmueller <martin@atzmueller.net>

References

1. Martin Atzmueller and Frank Puppe. SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery. Knowledge Discovery in Databases: PKDD 2006, LNAI 4213, pp. 6-17, Springer Verlag, 2006.
2. Martin Atzmueller and Florian Lemmerich. Fast Subgroup Discovery for Continuous Target Concepts. In: Foundations of Intelligent Systems, LNCS 5722, pp. 35-44, Springer Verlag, 2009.

3. Florian Lemmerich and Mathias Rohlfs and Martin Atzmueller. Fast Discovery of Relevant Subgroup Patterns. In: Proc. 23rd FLAIRS Conference, AAAI Press, 2010.

SDTaskConfig-class *Class “SDTaskConfig” — A Set of Configuration Settings*

Description

A Set of Configuration Settings for the Subgroup and Pattern Mining Algorithms

Objects from the Class

Objects are created by calls of the form `new("SDTaskConfig", ...)`.

Slots

qf: A quality function; one of: Binomial-Test `bin`, Chi-Square-Test `chi2`, Lift `lift`, Piatetsky-Shapiro `ps`, Gain `gain`, Relative Gain `relgain`, Weighted Relative Accuracy `wracc`.

method: A mining method; one of Beam-Search `beam`, BSD `bsd`, SD-Map `somap`, SD-Map enabling internal disjunctions `somap-dis`.

k: The maximum number (top-k) of patterns to discover.

minqual The minimal quality.

minsize The minimal size of a subgroup (minimal coverage of database records).

maxlen The maximal description length of a pattern, i.e., the maximal number of conjunctions.

nodefaults Ignore default values, i.e., do not include the respective first value of each attribute

relfilter Controls, whether irrelevant patterns are filtered during pattern mining; negatively impacts performance.

postfilter Controls, whether a post-processing filter is applied; one of: Minimum Improvement (Global) `min-improve-global`, checks the patterns against all possible generalizations, Minimum Improvement (Pattern Set) `min-improve-set`, checks the patterns against all their generalizations in the result set, Relevancy Filter `relevancy`, removes patterns that are strictly irrelevant, Significant Improvement (Global) `sig-improve-global`, removes patterns that do not significantly improve (0.05 level) w.r.t. all their possible generalizations, Significant Improvement (Set) `sig-improve-set`, removes patterns that do not significantly improve (0.05 level) w.r.t. all generalizations in the result set.

attributes The list of attributes to consider for mining. Either a vector of attribute names, or NULL, which includes all attributes.

See Also

[DiscoverSubgroups](#). [DiscoverSubgroupsByTask](#) [CreateSDTask](#)

`ToDataFrame`*Transforms patterns into a data frame*

Description

Transforms a list/vector of patterns into a data frame for inspection and analysis.

Usage

```
ToDataFrame(patterns, ndigits = 2)
```

Arguments

<code>patterns</code>	List/vector of patterns.
<code>ndigits</code>	Number of significant digits when printing floats (optional).

See Also

[DiscoverSubgroups](#).

Index

- *Topic **classes**
 - Pattern-class, 5
 - SDTaskConfig-class, 7
- *Topic **datasets**
 - credit.data, 3
- *Topic **package**
 - rsubgroup, 6
- *Topic **subgroup analysis**
 - ToDataFrame, 8
- *Topic **subgroup discovery**
 - DiscoverSubgroups, 4
- *Topic **subgroup task**
 - CreateSDTask, 3
 - DiscoverSubgroupsByTask, 5
- *Topic **target variable**
 - as.target, 2

as.target, 2, 4

CreateSDTask, 3, 4–7

credit.data, 3

DiscoverSubgroups, 2, 3, 4, 5–8

DiscoverSubgroupsByTask, 3, 4, 5, 6, 7

Pattern (Pattern-class), 5

Pattern-class, 5

rsubgroup, 6

SDTaskConfig, 3, 4

SDTaskConfig (SDTaskConfig-class), 7

SDTaskConfig-class, 7

ToDataFrame, 4, 5, 8