Package 'fpmoutliers'

November 22, 2017

Type Package

Title Frequent Pattern Mining Outliers

Version 0.1.0

Author Jaroslav Kuchar [aut, cre]

Maintainer Jaroslav Kuchar <jaroslav.kuchar@gmail.com>

URL https://github.com/jaroslav-kuchar/fpmoutliers

BugReports https://github.com/jaroslav-kuchar/fpmoutliers/issues

Description Algorithms for detection of outliers based on frequent pattern mining. Such algorithms follow the paradigm: if an instance contains more frequent patterns, it means that this data instance is unlikely to be an anomaly (He Zengyou, Xu Xiaofei, Huang Zhexue Joshua, Deng Shengchun (2005) <doi:10.2298/CSIS0501103H>).
The package implements a list of existing state of the art algorithms as well as other published approaches: FPI, WFPI, FPOF, FPCOF, LFPOF, MFPOF, WCFPOF and WFPOF.

License Apache License (== 2.0) | file LICENSE

LazyData TRUE

Encoding UTF-8

Depends R (>= 3.3.0)

Imports pmml, XML, Matrix, R.utils, arules (>= 1.5-4), foreach, doParallel, parallel, methods, pryr

RoxygenNote 6.0.1

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2017-11-22 09:25:49 UTC

R topics documented:

`build` *Automatic build of the anomaly detection model*

Description

An experimental implementataion that automatically builds an anomaly detection model

Usage

```
build(data, func = FPI, initial_support = 0.5, top_outlier_threshold = 3,
  iteration_timeout = 10)
```

Arguments

`data`	`data.frame` or `transactions` from `arules` with input data
`func`	function name of the method that will be use during the automatic build
`initial_support`	initial maximum support
`top_outlier_threshold`	number of top unique outliers as a stopping condition
`iteration_timeout`	timeout of one iteration

Value

model of outlier detection

Examples

```
# simple build with default parameters (FPI method)
library("fpmoutliers")
data("iris")
model <- fpmoutliers::build(iris[sample(nrow(iris), 5),])


# using other anomaly detection methods for the automatic build (e.g. LFPOF)
library("fpmoutliers")
data("iris")
model <- fpmoutliers::build(iris[sample(nrow(iris), 5),], func=LFPOF)
```

`describeInstance` *Explain the instance/outlier by a brief textual summary*

Description

Explain the instance/outlier by a brief textual summary

Usage

```
describeInstance(data, model, instanceIndex, topN = 10)
```

Arguments

`data`	data.frame with data describing all instances
`model`	outlier detection model
`instanceIndex`	index of the instance to visualize
`topN`	limit for a print of top matching frequent itemsets

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- FPI(dataFrame, minSupport = 0.001)
# sort data by the anomaly score
dataFrame <- dataFrame[order(model$scores, decreasing = TRUE),]
# instance with the highest anomaly score
describeInstance(dataFrame, model, 1)
# instance with the lowest anomaly score
describeInstance(dataFrame, model, nrow(dataFrame))
```

FPCOF *FPCOF algorithm*

Description

Algorithm proposed by: X. Tang, G. Li and G. Chen, "Fast Detecting Outliers over Online Data Streams," 2009 International Conference on Information Engineering and Computer Science, Wuhan, 2009, pp. 1-4.

Usage

```
FPCOF(data, minSupport = 0.3, mlen = 0, noCores = 1)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
noCores	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- FPCOF(dataFrame, minSupport = 0.001)
```

FPI *FPI - Frequent Pattern Isolation algorithm*

Description

Algorithm proposed by: J. Kuchar, V. Svatek: Spotlighting Anomalies using Frequent Patterns, Proceedings of the KDD 2017 Workshop on Anomaly Detection in Finance, Halifax, Nova Scotia, Canada, PMLR, 2017.

Usage

```
FPI(data, minSupport = 0.3, mlen = 0)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- FPI(dataFrame, minSupport = 0.001)
```

FPOF *FPOF - Frequent Pattern Outlier Factor algorithm*

Description

Algorithm proposed by: He, Z., Xu, X., Huang, J. Z., Deng, S.: FP-Outlier: Frequent Pattern Based Outlier Detection. Computer Science and Information Systems, Vol. 2, No. 1, 103-118. (2005)

Usage

```
FPOF(data, minSupport = 0.3, mlen = 0, noCores = 1)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
noCores	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- FPOF(dataFrame, minSupport = 0.001)
```

FPOFcontradictness *Frequent Pattern Outlier Factor*

Description

Frequent Pattern Outlier Factor

Usage

```
FPOFcontradictness(dataFrame, anIndex, minSupport = 0.3, mlen = 0, k = 10)
```

Arguments

dataFrame	data.frame with input data
anIndex	anomaly index
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
k	top-k contradictness

Value

vector with outlier scores

generatePMML *PMML conversion - PMML representation of outliers*

Description

Kuchar, Jaroslav et al. "Outlier (Anomaly) Detection Modelling in PMML." RuleML+RR (2017).(http://ceur-ws.org/Vol-1875/paper9.pdf)

Usage

```
generatePMML(model, dataFrame = NULL, topN = NULL)
```

Arguments

model	outlier model
dataFrame	frame for labeling
topN	limit number of outliers in the output

Value

pmml model

Examples

```
## Not run:
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- FPI(dataFrame, minSupport = 0.001)
generatePMML(model, dataFrame)

## End(Not run)
```

LFPOF *LFPOF algorithm*

Description

Algorithm proposed by: W. Zhang, J. Wu and J. Yu, "An Improved Method of Outlier Detection Based on Frequent Pattern," Information Engineering (ICIE), 2010 WASE International Conference on, Beidaihe, Hebei, 2010, pp. 3-6.

Usage

```
LFPOF(data, minSupport = 0.3, mlen = 0, noCores = 1)
```

Arguments

`data`	`data.frame` or `transactions` from `arules` with input data
`minSupport`	minimum support for FPM
`mlen`	maximum length of frequent itemsets
`noCores`	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- LFPOF(dataFrame, minSupport = 0.001)
```

MFPOF *MFPOF algorithm*

Description

Algorithm proposed by: Feng Lin, Wang Le, Jin Bo - Research on Maximal Frequent Pattern Outlier Factor for Online HighDimensional Time-Series Outlier Detection. Journal of Convergence Information Technology 5(10):66-71. December 2010.

Usage

```
MFPOF(data, minSupport = 0.3, mlen = 0, noCores = 1)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
noCores	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- MFPOF(dataFrame, minSupport = 0.001)
```

parsePMML *PMML parser*

Description

The parser parses the proposed PMML for the outlier detection model and build its object representation.

Usage

```
parsePMML(fileName)
```

Arguments

fileName	xml file name

Value

list model

Examples

```
## Not run:
library("fpmoutliers")
model <- parsePMML("od-pmml.xml")

## End(Not run)
```

`plotRestrictedBarplot` *Plot of a restricted barplot with a highlighted selected value*

Description

Plot of a restricted barplot with a highlighted selected value

Usage

```
plotRestrictedBarplot(data, selectedValue, bars = 10, title = "")
```

Arguments

`data`	vector with all values
`selectedValue`	value that will be highlighted
`bars`	max number of bars to plot, only top lowest and top highest frequencies will be presented
`title`	main title of the plot

Examples

```
library("fpmoutliers")
data <- sample(1:1000, 1000, replace = TRUE)
selectedValue <- data[length(data)/2]
plotRestrictedBarplot(data, selectedValue, bars = 15,
     title=paste("Experimental=",selectedValue,sep=""))
```

visualizeInstance *Visualization of a data instance using a set of barplots*

Description

Visualization of a data instance using a set of barplots

Usage

```
visualizeInstance(data, instanceIndex, bars = 10)
```

Arguments

data	data.frame with data describing all instances
instanceIndex	index of the instance to visualize
bars	max number of bars to plot, only top lowest and top highest frequencies will be presented

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- FPI(dataFrame, minSupport = 0.001)
# sort data by the anomaly score
dataFrame <- dataFrame[order(model$scores, decreasing = TRUE),]
visualizeInstance(dataFrame, 1) # instance with the highest anomaly score
visualizeInstance(dataFrame, nrow(dataFrame)) # instance with the lowest anomaly score
```

WCFPOF *WCFPOF algorithm*

Description

Algorithm proposed by: Jiadong Ren, Qunhui Wu, Jiadong Ren, Changzhen Hu, Kunsheng Wang - An Approach for Analyzing Infrequent Software Faults Based on Outlier Detection

Usage

```
WCFPOF(data, minSupport = 0.3, mlen = 0, noCores = 1)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
noCores	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- WCFPOF(dataFrame, minSupport = 0.001)
```

WFPI *WFPI - Weighted Frequent Pattern Isolation algorithm*

Description

Algorithm proposed by: J. Kuchar, V. Svatek: Spotlighting Anomalies using Frequent Patterns, Proceedings of the KDD 2017 Workshop on Anomaly Detection in Finance, Halifax, Nova Scotia, Canada, PMLR, 2017.

Usage

```
WFPI(data, minSupport = 0.3, mlen = 0, preferredColumn = "",
  preference = 1, noCores = 1)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
preferredColumn	column name that is preferred
preference	numeric value that multiplies the score
noCores	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- WFPI(dataFrame, minSupport = 0.001, preferredColumn="Car", preference=10)
```

WFPOF *WFPOF algorithm*

Description

Algorithm proposed by: ZHOU Xiao-Yun+, SUN Zhi-Hui, ZHANG Bai-Li, YANG Yi-Dong - A Fast Outlier Detection Algorithm for High Dimensional Categorical Data Streams. Journal of Software 18(4). April 2007.

Usage

```
WFPOF(data, minSupport = 0.3, mlen = 0, noCores = 1)
```

Arguments

data	data.frame or transactions from arules with input data
minSupport	minimum support for FPM
mlen	maximum length of frequent itemsets
noCores	number of cores for parallel computation

Value

model output (list) with all results including outlier scores

Examples

```
library("fpmoutliers")
dataFrame <- read.csv(
     system.file("extdata", "fp-outlier-customer-data.csv", package = "fpmoutliers"))
model <- WFPOF(dataFrame, minSupport = 0.001)
```

Index