Package 'OutlierDM'

July 2, 2014

Title Outlier detection for replicated high-throughput data

Date 2014-03-22

Version 1.0-0

Description Detecting outlying observations for replicated high-throughput high-dimensional data based on a MA plot with quantile regression approach.

Depends R (>= 3.0.0), quantreg, MatrixModels, outliers, pcaPP

Imports methods, graphics

Maintainer Soo-Heang Eo <eo.sooheang@gmail.com>

License GPL-3

Author Soo-Heang Eo [aut, cre],HyungJun Cho [aut]

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-23 07:03:41

R topics documented:

`OutlierDM-package` *Functions for detecting outlying observations in (multiple) replicated high-throughput data*

Description

This package provides outlier detection algorithms for multiple replicated high-throughput high-dimensional data, especially in the field of mass spectrometry.

Details

Package:	OutlierDM
Type:	Package
Version:	1.0-0
Date:	2014-03-22
License:	GPL version 3
LazyLoad:	no

Author(s)

Soo-Heang Eo <eo.sooheang@gmail.com> and HyungJun Cho <hj4cho@korea.ac.kr>

Maintainer: Soo-Heang Eo <eo.sooheang@gmail.com>

References

S-H Eo, D Pak, J Choi, H Cho (2012). Outlier Detection for Multiplicative High-throughput Data. *BMC Research Notes*, **5**, 1–6.

Cho et al. (2008). OutlierD: an R package for outlier detection using quantile regression on mass spectrometry data. *Bioinformatics*, **24**(6), 882–884.

Min et al. (2007). Ultrahigh-pressure dual online solid phase extraction/capillary reverse-phase liquid chromatography/tandem mass spectrometry (DO-SPE/cRPLC/MS/MS): A versatile separation platform for high-throughput and highly sensitive proteomic analyses. *ELECTROPHORESIS*, **28**, 1012–1021.

See Also

`odm`, `odm.control`, `quantreg`

Examples

```
## Not run:
data(lcms3)
```

```
## Fit projection approaches
fit.proj.const <- odm(lcms3, method="constant")
fit.proj.linear <- odm(lcms3, method="linear")
fit.proj.nonlin <- odm(lcms3, method="nonlin")
fit.proj.nonpara <- odm(lcms3, method="nonpar", lbda = 1)

par(mfrow = c(2,2))
plot(fit.proj.const, main = "Constant")
plot(fit.proj.linear, main = "Linear")
plot(fit.proj.nonlin, main = "NonLinear")
plot(fit.proj.nonpara, main = "Nonparametric")

## Fit pairwise OutlierD algorithm
fit0 <- odm(lcms3, type = "pair")
plot(fit0)

## End(Not run)
```

`lcms3` *LC/MS/MS data*

Description

This data set consists of three-replicated LC/MS/MS data obtained from laboratory of gaseous ion chemistry in department of chemistry, Korea University. The data is already converted into log-2 transformation.

Usage

```
data(lcms3)
```

Format

a matrix for LCMS data, rows=peptides, columns=samples

Source

Min et al. (2007). Ultrahigh-pressure dual online solid phase extraction/capillary reverse-phase liquid chromatography/tandem mass spectrometry (DO-SPE/cRPLC/MS/MS): A versatile separation platform for high-throughput and highly sensitive proteomic analyses. *ELECTROPHORESIS*, **28**, 1012–1021.

Examples

```
data(lcms3)
```

odm *Outlier dectection for (multiple) replicated high-throughput data*

Description

This function is desinged to detect outlying observations, or peptides, based on a MA plot with boxplot by quantile regression methods.

Usage

```
   odm(x, k=1.5, method= c("nonlin", "constant", "linear", "nonpar"),
type = c("proj", "diff", "pair", "grubbs", "dixon"), ...)
```

Arguments

x	data vectors or matrices. These can be given as named arguments. If the number of predictors is 2, x1 describes one n-by-1 vector for data and x2 describes the other n-by-1 vector for data (n= number of peptides, proteins, or genes)
k	a tuning parameter used in Q1-k*IQR and Q3+k*IQR, IQR=Q3-Q1. Default value is 1.5
method	Selet a type of quantile regression methods used in an outlier detection algorithm. Use one of "constant", "linear", "nonlin", and "nonpar"m which mean constant, linear, non-linear, and nonparametric quantile regression in order. For more details, see the quantreg package.
type	input parameter for multiplicative detection algorithm. Use one of "dixon", "grubbs", "pair", "diff", and "proj". dixon: Dixon's Q-test grubbs: Grubbs' test pair: OutlierD algorithm (Cho et al., 2008) diff: Difference approach for multiplicative highthroughput data proj: Outlier Detection using projections for multiplicative experiments (Eo et al., 2012)
...	tuning parameters used in odm.control(). See odm.control.

Details

Outlier detection is an important issue in high-throughput data generated from various biological or chemical experiments such as mass spectrometry and microarrays because elaborate pre-processing is essential for successful analysis and outlier detection is one of pre-processing steps. The heterogeneity of variability and low replication are often obstables for successful analysis including outlier detection. An outlier detection algorithm using quantile regression was developed to solve the problems; however, it can be used only for duplicate experiments. We here proposed an outlier detection algorithm using projection for multiplicative experiments.

Value

`call:`	evaluated function call
`raw.data:`	data to be used in the fitted model
`res:`	a data.frame including the information about the fitted model. It consists of several columns including `outlier, M, A, Q3, Q1, UB` and `LB`.
`x.pair:`	Object of class `"list"` ~~
`k:`	a scalar parameter for constructing boxplot used in the fitted models
`n.outliers:`	a scalar value that denotes the number of outliers to be detected by the fitted model.
`method:`	the type of method used in the fitted model
`type:`	the type of algorithm used in the fitted model
`contrl.para:`	a list including information about tuning parameters

References

S-H Eo, D Pak, J Choi, H Cho (2012). Outlier Detection for Multiplicative High-throughput Data. *BMC Research Notes*, **5**, 1–6.

See Also

`OutlierDM-package` to show general information about the package
`OutlierDM-class` to provide information about the "OutlierDM" class
`odm.control` to control tuning parameters

Examples

```
## Not run:
data(lcms3)

## Fit projection approaches
fit.proj.const <- odm(lcms3, method="constant")
fit.proj.linear <- odm(lcms3, method="linear")
fit.proj.nonlin <- odm(lcms3, method="nonlin")
fit.proj.nonpara <- odm(lcms3, method="nonpar", lbda = 1)

par(mfrow = c(2,2))
plot(fit.proj.const, main = "Constant")
plot(fit.proj.linear, main = "Linear")
plot(fit.proj.nonlin, main = "NonLinear")
plot(fit.proj.nonpara, main = "Nonparametric")

## Fit pairwise OutlierD algorithm
fit0 <- odm(lcms3, type = "pair")
plot(fit0)

## End(Not run)
```

odm.control *Control tuning parameters for "OutlierDM" object*

Description

various parameters that control aspects of the "OutlierDM" object

Usage

```
      odm.control(pair.cre = 1, dist.mthd = "median",
       Lower = .25, Upper = .75, trans = "log2",
       centering = TRUE, projection.type = "PCA", lbda = 1,
nonlin.method = "L-BFGS-B", nonlin.SS = "AsymOff",
nonlin.Frank = c(2, -8, 0, 1), ncl = 2 )
```

Arguments

pair.cre	a scalar parameter to specify the minimum number of pairs, used in "type = pair".
dist.mthd	a distance parameter used in "type = diff". you can choose one of "median", "mean" or and so on.
Lower	a criterion for lower quantile value used to construct boxplot
Upper	a criterion for upper quantile value used to construct boxplot
trans	a parameter for a logarithm and exponential transformation. If a log 2 transformation is needed, set "trans = log2". If no trasformation is needed, set "trans = FALSE".
centering	a logical parameter for the status of centering. If "centering = TRUE", data are centered by its column means.
projection.type	a parameter to determine a tyoe of projection methods. Choose one of "naive", "pca", and "robust".
lbda	a criterion about lambda used for nonlinear quantile regression.
nonlin.method	a parameter to determine a type of methods used for nonlinear quantile regression. choose one of "L-BFGS-B" and "BFGS". Default is "L-BFGS-B".
nonlin.SS	a parameter to determine a type of structure used for nonlinear quantile regression. choose one of "Frank", "Self", "Asym" and "AsymOff". Default is "AsymOff", Asymptotic Regression Model with an Offset.
nonlin.Frank	a structure parameter used for Frank copula model. Gain c(df, delta, mu, sigma) in the Frank copula formula
ncl	A parameter to determine the number of cores used in parallel computing. A default value is 2.

See Also

odm

OutlierDM-class *Class* "OutlierDM"

Description

A class for outlier detection algorithms for high-throughput high-dimensional data.

Objects from the Class

Objects can be created by calls of the form `new("OutlierDM", ...)`. See following information about slots.

Slots

`call`: evaluated function call

`raw.data`: data to be used in the fitted model

`res`: a data.frame including the information about the fitted model. It consists of several columns including `outlier, M, A, Q3, Q1, UB` and `LB`.

`x.pair`: Object of class `"list"` ~~

`k`: a scalar parameter for constructing boxplot used in the fitted models

`n.outliers`: a scalar value that denotes the number of outliers to be detected by the fitted model.

`method`: the type of method used in the fitted model

`type`: the type of algorithm used in the fitted model

`contrl.para`: a list including information about tuning parameters

Methods

show `signature(object = "OutlierDM")`: Same as the show method without the optional argumetns

plot `signature(x = "OutlierDM", y = "missing")`: Plot an object.

See Also

odm

Examples

```
showClass("OutlierDM")
```

plot *a plot-method for a "OutlierDM" object*

Description

This function provides a MA scatter plot with quantile regression based boxlplot.

Usage

```
    ## S4 method for signature 'OutlierDM'
plot(x, y = NA, pch = 20, cex = 0.5, xlab ="A", ylab = "M",
legend.use = TRUE, ...)
```

Arguments

x	fitted model object of class odm.
y	the "y" argument is not used in the plot-method for "OutlierDM" object.
pch	a vector of plotting characters or symbols: see plot.defalut.
cex	See plot.defalut.
xlab	label X
ylab	label Y
legend.use	logical option for using legend box
...	plot.default arguments

Details

This function is a method for the generic function plot for the S4 class OutlierDM. It can be invoked by calling print for an object of the appropriate class, or directly by calling plot.OutlierDM regardless of the class of the object.

See Also

odm

Index