Package 'EnergyOnlineCPM'

January 20, 2025

Type Package

Title Distribution Free Multivariate Control Chart Based on Energy
Test

Version 1.0

Date 2017-09-30

Author Yafei Xu

Maintainer Yafei Xu <yafei.xu@hotmail.de>

Depends parallel, energy, MASS

Description

Provides a function for distribution free control chart based on the change point model, for multivariate statistical process control. The main constituent of the chart is the energy test that focuses on the discrepancy between empirical characteristic functions of two random vectors. This new control chart highlights in three aspects. Firstly, it is distribution free, requiring no knowledge of the random processes. Secondly, this control chart can monitor mean and variance simultaneously. Thirdly it is devised for multivariate time series which is more practical in real data application. Fourthly, it is designed for online detection (Phase II), which is central for real time surveillance of stream data. For more information please refer to O. Okhrin and Y.F. Xu (2017) <https://github.com/YafeiXu/working_paper/raw/master/CPM102.pdf>.

URL https://sites.google.com/site/EnergyOnlineCPM/

License GPL (>= 3)

NeedsCompilation no

Repository CRAN

Date/Publication 2017-10-02 11:37:01 UTC

Contents

1

`EnergyOnlineCPM-package`
EnergyOnlineCPM Package

Description

EnergyOnlineCPM provides users a new function for nonparametric Phase II MSPC.

Details

This package provides users a new function for nonparametric Phase II multiple multivariate change points detection.

Author(s)

Yafei Xu <yafei.xu@hotmail.de>

References

[1] Szekely, G. J. & Rizzo, M. L. (2004). Testing for equal distributions in high dimension, InterStat.

[2] Hawkins, D. M., Qiu, P. & Kang, C. W. (2003). The changepoint model for statistical process control, Journal of Quality Technology.

[3] Xu, Y. F. (2017). Reference manual: An R package "EnergyOnlineCPM".

see URL: https://sites.google.com/site/EnergyOnlineCPM/

`maxEnergyCPMv`   *Nonparametric Multivariate Control Chart based on Energy Test*

Description

This R function centers on nonparametric Phase II multiple change points detection for high dimensional time series. Three highlights are included in the function. Firstly, the new model is nonparametric which does not require any distributional pre-knowledge about the process. The test is based on the maximum energy statistic (see [1] Gabor J. Szekely and Maria L. Rizzo 2004, Testing for Equal Distributions in High Dimension) and permutation samples. Secondly, the model is a Phase II change point model (see [2]) which is used for online detection of stream data not for batch data. Phase II set-up has practical meaning in time series change detection. Thirdly, it is concentrated on high dimensional data, i.e. multivariate context. An important remark is that the data used in this function must be independent, i.e. every row in the N*d matrix must be an independent observation. If your data set contains not-independent observations then you need to handle the data using some filter functions, e.g. multivariate time series model to filter out the residuals which are theoretically independent.

Usage

```
maxEnergyCPMv(data1, wNr, permNr, alpha)
```

Arguments

data1	an N*d matrix, N is the number of observations and d the dimensions.
wNr	a scalar of warm-up.
permNr	a scalar of times of permutation.
alpha	a scalar of significant level

Details

The function returns ONLY ONE vector containing even number components, where the first half stands for detection time vector and the rest half stands for the vector of change time locations.

Value

result	a vector of locations of detection time in the first half, locations of change time in the second half.

Author(s)

Yafei Xu <yafei.xu@hotmail.de>

References

[1] Szekely, G. J. & Rizzo, M. L. (2004). Testing for equal distributions in high dimension, InterStat.

[2] Hawkins, D. M., Qiu, P. & Kang, C. W. (2003). The changepoint model for statistical process control, Journal of Quality Technology.

[3] Xu, Y. F. (2017). Reference manual: An R package "EnergyOnlineCPM".

see URL: https://sites.google.com/site/EnergyOnlineCPM/

Examples

```
# simulate 300 length time series
simNr=300

# simulate 300 length 5 dimensonal standard Gaussian series
Sigma2 <- matrix(c(1,0,0,0,0,  0,1,0,0,0,   0,0,1,0,0,  0,0,0,1,0,  0,0,0,0,1),5,5)
Mean2=rep(1,5)
sim2=(mvrnorm(n = simNr, Mean2, Sigma2))

# simulate 300 length 5 dimensonal standard Gaussian series
Sigma3 <- matrix(c(1,0,0,0,0,  0,1,0,0,0,   0,0,1,0,0,  0,0,0,1,0,  0,0,0,0,1),5,5)
Mean3=rep(0,5)
sim3=(mvrnorm(n = simNr, Mean3, Sigma3))

# construct a data set of length equal to 35.
# first 20 points are from standard Gaussian.
```

```
# second 15 points from a Gaussian with a mean shift with 555.
data1=sim6=rbind(sim2[1:20,],(sim3+555)[1:15,])

# set warm-up number as 20, permutation 200 times, significant level 0.005
wNr=20
permNr=200
alpha=1/200
maxEnergyCPMv(data1,wNr,permNr,alpha)
```

Index