Package 'VIFCP'

August 31, 2016

Version 1.2

Title Detecting Change-Points via VIFCP Method

Author Xiaoping Shi, Xiangsheng Wang, Dongwei Wei, Yuehua Wu

Maintainer Dongwei Wei <weidw@mathstat.yorku.ca>

Depends R (>= 3.1.3)

License LGPL-3

Description Contains a function to support the following paper:
Xiaoping Shi, Xiang-Sheng Wang, Dongwei Wei, Yuehua Wu (2016), <DOI:10.1007/s00180-015-0587-5>,
A sequential multiple change-point detection procedure via VIF regression,
Computational Statistics, 31(2): 671-691.

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-08-31 20:46:19

R topics documented:

`vif.cp` *Detect the Location of Change-Points via VIFCP Method*

Description

A modified variance inflation factor (VIF) regression algorithm is used to perform the variable selection sequentially in segment order. We use the idea of VIF to detect change points in order to increase the speed of computation.

Usage

```
vif.cp(data, l, siglev=0.05)
```

Arguments

data	The original dataset that may contains change points for detection.
l	The length of the partition. See reference for more details.
siglev	The level of significance, default=0.05.

Value

The return value is the location of change points. If the return is 0, it means there is no change point in the dataset.

Author(s)

Xiaoping Shi, Xiangsheng Wang, Dongwei Wei, Yuehua Wu

Maintainer: Dongwei Wei (weidw@mathstat.yorku.ca)

References

Xiaoping Shi, Xiang-Sheng Wang, Dongwei Wei, Yuehua Wu. (2016). A sequential multiple changepoint detection procedure via VIF regression. Computational Statistics. 31(2): 671-691.

Examples

```
###example 1: simple case
y<-c(rnorm(100,0,0.2),rnorm(200,0.4,0.2),rnorm(100,0.9,0.2))
vif.cp(y,50,0.05)

###example 2: Reference paper's simulation study
n <- 2000
location_cp=n*c(.162,.31,.551,.693,.805)
location_cp
number_cp = length(location_cp)
beta = c(0,0.3,0.7,0.2,-0.4,0.3) #beta: parameters in different blocks

newlocation <- c(0,location_cp,n)
y_true <- NULL                  ##generate the mean of each blocks
for(i in 1:(number_cp+1)){
   y_true <- c(y_true, rep(beta[i], newlocation[i+1]-newlocation[i]))
}

y_error <- y_true+rnorm(n,0,0.2) ##add white noise with sd=0.2
vif.cp(y_error, 100, 0.05)

y_error <- y_true+rnorm(n,0,0.3) ##add white noise with sd=0.3
vif.cp(y_error, 100, 0.05)

y_error <-y_true+rnorm(n,0,0.4) ##add white noise with sd=0.4
vif.cp(y_error, 100, 0.05)


###example 3: Re-present the results of Table 1-3
```

```
rm(list=ls()) #remove or clear all variables
library(VIFCP)
set.seed(3)
###function to calculate the number of successful detection
###In the paper, we use distance=5
count<-function(A,B,distance){
        n1<-length(A)
        n2<-length(B)
        result<-0
        for(i in 1:n2){
            result<-result+as.numeric(sum(abs(A-B[i])<=distance)>=1)
        }
        return(result)
}

##save results for S1
error1.1<-matrix(0,5,3)
rownames(error1.1)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp")
colnames(error1.1)<-c("l=100,sigma=0.2","l=100,sigma=0.3","l=100,sigma=0.4")

error1.2<-matrix(0,5,3)
rownames(error1.2)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp")
colnames(error1.2)<-c("l=80,sigma=0.2","l=80,sigma=0.3","l=80,sigma=0.4")


cpnumber.R1<-matrix(0,2,3)
rownames(cpnumber.R1)<-c("l=100","l=80")
colnames(cpnumber.R1)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

corr.det1<-matrix(0,2,3)
rownames(corr.det1)<-c("l=100","l=80")
colnames(corr.det1)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

ERT.S1<-matrix(0,2,3)
rownames(ERT.S1)<-c("l=100","l=80")
colnames(ERT.S1)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

###save results for S2
error2.1<-matrix(0,5,3)
rownames(error2.1)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp")
colnames(error2.1)<-c("l=100,sigma=0.2","l=100,sigma=0.3","l=100,sigma=0.4")

error2.2<-matrix(0,5,3)
rownames(error2.2)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp")
colnames(error2.2)<-c("l=80,sigma=0.2","l=80,sigma=0.3","l=80,sigma=0.4")


cpnumber.R2<-matrix(0,2,3)
rownames(cpnumber.R2)<-c("l=100","l=80")
colnames(cpnumber.R2)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

corr.det2<-matrix(0,2,3)
rownames(corr.det2)<-c("l=100","l=80")
```

```
colnames(corr.det2)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

ERT.S2<-matrix(0,2,3)
rownames(ERT.S2)<-c("l=100","l=80")
colnames(ERT.S2)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

###save results for S3
error3.1<-matrix(0,5,3)
rownames(error3.1)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp")
colnames(error3.1)<-c("l=100,sigma=0.2","l=100,sigma=0.3","l=100,sigma=0.4")

error3.2<-matrix(0,5,3)
rownames(error3.2)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp")
colnames(error3.2)<-c("l=80,sigma=0.2","l=80,sigma=0.3","l=80,sigma=0.4")


cpnumber.R3<-matrix(0,2,3)
rownames(cpnumber.R3)<-c("l=100","l=80")
colnames(cpnumber.R3)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

corr.det3<-matrix(0,2,3)
rownames(corr.det3)<-c("l=100","l=80")
colnames(corr.det3)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

ERT.S3<-matrix(0,2,3)
rownames(ERT.S3)<-c("l=100","l=80")
colnames(ERT.S3)<-c("sigma=0.2","sigma=0.3","sigma=0.4")

sigma<-c(0.2,0.3,0.4)

for(dn in 1:3){

    n<-2000
    location_cp=n*c(.162,.31,.551,.693,.805)

    number_cp=length(location_cp)
    beta=c(0,0.3,0.7,0.2,-0.2,0.3) #beta: parameters in different blocks
    newlocation<-c(0,location_cp,n)
    y_true<-NULL ##generate the mean of each blocks
    for(i in 1:(number_cp+1)){
          y_true<-c(y_true, rep(beta[i],newlocation[i+1]-newlocation[i]))
    }
  for(loop in 1:1000){
    error.term<-rnorm(n,0,sigma[dn])


    #generate data for S1
    scenario1<-y_true+error.term

    #generate data for S2
    outlier1<-sample(1:n,5)
    error.term1<-error.term
    error.term1[outlier1]<-5+error.term1[outlier1]
```

```
scenario2<-y_true+error.term1

#generate data for S3
outlier2<-sample(1:n,10)
error.term2<-error.term
error.term2[outlier2]<-5+error.term2[outlier2]
scenario3<-y_true+error.term2

##for S1; l=100
time1<-proc.time()[3]
cp.vif11<-vif.cp(scenario1,l=100,siglev=0.05)
ERT.S1[1,dn]<-ERT.S1[1,dn]+proc.time()[3]-time1

for(kk in 1:5){
     error1.1[kk,dn]<-error1.1[kk,dn]+count(cp.vif11,location_cp[kk],distance=5)
}
if(length(cp.vif11)==number_cp){
             cpnumber.R1[1,dn]<-cpnumber.R1[1,dn]+1
             temp<-count(cp.vif11, location_cp,distance=5)
             corr.det1[1,dn]<-corr.det1[1,dn]+as.numeric(temp==number_cp)
}

##for S1; l=80
time1<-proc.time()[3]
cp.vif12<-vif.cp(scenario1,l=80,siglev=0.05)
ERT.S1[2,dn]<-ERT.S1[2,dn]+proc.time()[3]-time1

for(kk in 1:5){
     error1.2[kk,dn]<-error1.2[kk,dn]+count(cp.vif12,location_cp[kk],distance=5)
}
if(length(cp.vif12)==number_cp){
             cpnumber.R1[2,dn]<-cpnumber.R1[2,dn]+1
             temp<-count(cp.vif12, location_cp,distance=5)
             corr.det1[2,dn]<-corr.det1[2,dn]+as.numeric(temp==number_cp)
}

########for S2; l=100
time1<-proc.time()[3]
cp.vif21<-vif.cp(scenario2,l=100,siglev=0.05)
ERT.S2[1,dn]<-ERT.S2[1,dn]+proc.time()[3]-time1

for(kk in 1:5){
     error2.1[kk,dn]<-error2.1[kk,dn]+count(cp.vif21,location_cp[kk],distance=5)
}
if(length(cp.vif21)==number_cp){
             cpnumber.R2[1,dn]<-cpnumber.R2[1,dn]+1
             temp<-count(cp.vif21, location_cp,distance=5)
             corr.det2[1,dn]<-corr.det2[1,dn]+as.numeric(temp==number_cp)
}

##for S2; l=80
time1<-proc.time()[3]
cp.vif22<-vif.cp(scenario2,l=80,siglev=0.05)
```

```
    ERT.S2[2,dn]<-ERT.S2[2,dn]+proc.time()[3]-time1

    for(kk in 1:5){
         error2.2[kk,dn]<-error2.2[kk,dn]+count(cp.vif22,location_cp[kk],distance=5)
    }
    if(length(cp.vif22)==number_cp){
                 cpnumber.R2[2,dn]<-cpnumber.R2[2,dn]+1
                 temp<-count(cp.vif22, location_cp,distance=5)
                 corr.det2[2,dn]<-corr.det2[2,dn]+as.numeric(temp==number_cp)
    }

    ############for S3; l=100
        time1<-proc.time()[3]
    cp.vif31<-vif.cp(scenario3,l=100,siglev=0.05)
    ERT.S3[1,dn]<-ERT.S3[1,dn]+proc.time()[3]-time1

    for(kk in 1:5){
         error3.1[kk,dn]<-error3.1[kk,dn]+count(cp.vif31,location_cp[kk],distance=5)
    }
    if(length(cp.vif31)==number_cp){
                 cpnumber.R3[1,dn]<-cpnumber.R3[1,dn]+1
                 temp<-count(cp.vif31, location_cp,distance=5)
                 corr.det3[1,dn]<-corr.det3[1,dn]+as.numeric(temp==number_cp)
    }

    ##for S3; l=80
    time1<-proc.time()[3]
    cp.vif32<-vif.cp(scenario3,l=80,siglev=0.05)
    ERT.S3[2,dn]<-ERT.S3[2,dn]+proc.time()[3]-time1

    for(kk in 1:5){
         error3.2[kk,dn]<-error3.2[kk,dn]+count(cp.vif32,location_cp[kk],distance=5)
    }
    if(length(cp.vif32)==number_cp){
                 cpnumber.R3[2,dn]<-cpnumber.R3[2,dn]+1
                 temp<-count(cp.vif32, location_cp,distance=5)
                 corr.det3[2,dn]<-corr.det3[2,dn]+as.numeric(temp==number_cp)
    }
  }
}


####Build Table 1
Table1<-matrix(0,8,6)
rownames(Table1)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp",
     "cpnumber.R","ALLCP","ERT.S")
colnames(Table1)<-c("l=100,sigma=0.2","l=100,sigma=0.3","l=100,sigma=0.4",
 "l=80,sigma=0.2","l=80,sigma=0.3","l=80,sigma=0.4")

 Table1[1:5,1:3]<-error1.1
 Table1[1:5,4:6]<-error1.2
 Table1[6,]<-c(cpnumber.R1[1,],cpnumber.R1[2,])
 Table1[7,]<-c(corr.det1[1,]/cpnumber.R1[1,],corr.det1[1,]/cpnumber.R1[1,])
```

```
 Table1[8,]<-c(ERT.S1[1,],ERT.S1[2,])
 cat("Table 1: Results for Scenario 1\n")
 Table1


####Build Table 2
 Table2<-matrix(0,8,6)
rownames(Table2)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp",
     "cpnumber.R","ALLCP","ERT.S")
colnames(Table2)<-c("l=100,sigma=0.2","l=100,sigma=0.3","l=100,sigma=0.4",
 "l=80,sigma=0.2","l=80,sigma=0.3","l=80,sigma=0.4")

 Table2[1:5,1:3]<-error2.1
 Table2[1:5,4:6]<-error2.2
 Table2[6,]<-c(cpnumber.R2[1,],cpnumber.R2[2,])
 Table2[7,]<-c(corr.det2[1,]/cpnumber.R2[1,],corr.det2[1,]/cpnumber.R2[1,])
 Table2[8,]<-c(ERT.S2[1,],ERT.S2[2,])
 cat("Table 2: Results for Scenario 2\n")
 Table2

##Build Table 3
  Table3<-matrix(0,8,6)
  rownames(Table3)<-c("1st cp","2nd cp", "3rd cp", "4th cp", "5th cp",
      "cpnumber.R","ALLCP","ERT.S")
  colnames(Table3)<-c("l=100,sigma=0.2","l=100,sigma=0.3","l=100,sigma=0.4",
      "l=80,sigma=0.2","l=80,sigma=0.3","l=80,sigma=0.4")

 Table3[1:5,1:3]<-error3.1
 Table3[1:5,4:6]<-error3.2
 Table3[6,]<-c(cpnumber.R3[1,],cpnumber.R3[2,])
 Table3[7,]<-c(corr.det3[1,]/cpnumber.R3[1,],corr.det3[1,]/cpnumber.R3[1,])
 Table3[8,]<-c(ERT.S3[1,],ERT.S3[2,])
 cat("Table 3: Results for Scenario 3\n")
 Table3
```

Index