experiment.R

matt

Sun Jun 11 14:05:54 2017

```
library(changepoint)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Successfully loaded changepoint package version 2.2.2
##  NOTE: Predefined penalty values changed in version 2.2.  Previous penalty values with a postfix 1 i
```

```
library(ROCR)
```

```
## Loading required package: gplots
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

segment data set based on changepoints

```
SegmentDataset <- function(dataset, points) {

}
```

calculate F1 score based on ground truth

```
CalculateF1 <- function(dataset, points, truth) {
  algo.predictions <- numeric(length(dataset))
  ground.truth <- numeric(length(dataset))

  for(i in truth) {
    ground.truth[i] <- 1
  }

  for(i in points) {
    algo.predictions[i] <- 1
  }

  pred.obj <- prediction(algo.predictions, ground.truth)
  perf <- performance(pred.obj, "f")
  plot(perf)
}
```

main experiment method

```
RunExperiment <- function(input, daily = TRUE) {
  dataset <- ProcessData(input, daily)
  groundtruth <- GetGroundTruth(input)

  #get indices for ground truth
  library(foreach)
  groundtruth.indexed <-
    foreach(i = groundtruth) %do% which(dataset$Date == i)

  # set up the plot area
  par(mfrow = c(3, 3))

  # run the experiments!

  # Mean PELT
  mean.pelt <- cpt.mean(
    dataset$Freq,
    method = "PELT",
    penalty = "Hannan-Quinn",
    minseglen = 5
  )

  plot(mean.pelt, main = "Mean w/PELT", ylab = "Postings")
  PlotGroundTruth(groundtruth.indexed)
  CalculateF1(mean.pelt@data.set, mean.pelt@cpts, groundtruth.indexed)

  # Mean SegNeigh
  mean.segneigh <- cpt.mean(dataset$Freq,
                            method = "SegNeigh",
                            penalty = "Hannan-Quinn",
                            Q = 5)

  plot(mean.segneigh, main = "Mean w/SegNeigh", ylab = "Postings")
  PlotGroundTruth(groundtruth.indexed)


  # Mean BinSeg
  mean.binseg <- cpt.mean(
    dataset$Freq,
    method = "BinSeg",
    test.stat = "CUSUM",
    penalty = "Hannan-Quinn",
    minseglen = 5,
    Q = 5
  )

  plot(mean.binseg, main = "Mean w/BinSeg", ylab = "Postings")
  PlotGroundTruth(groundtruth.indexed)


  # Var PELT
  var.pelt <- cpt.var(
    dataset$Freq,
```

```
  method = "PELT",
  penalty = "Hannan-Quinn",
  minseglen = 5
)

plot(var.pelt, main = "Variance w/PELT", ylab = "Postings")
PlotGroundTruth(groundtruth.indexed)


# Var SegNeigh
var.segneigh <- cpt.var(dataset$Freq,
                        method = "SegNeigh",
                        penalty = "Hannan-Quinn",
                        Q = 5)

plot(var.segneigh, main = "Variance w/SegNeigh", ylab = "Postings")
PlotGroundTruth(groundtruth.indexed)

# Var BinSeg
var.binseg <- cpt.var(dataset$Freq,
                      method = "BinSeg",
                      #test.stat = "CSS",
                      penalty = "Hannan-Quinn",
                      Q = 5)

plot(var.binseg, main = "Variance w/BinSeg", ylab = "Postings")
PlotGroundTruth(groundtruth.indexed)


# MeanVar PELT
meanvar.pelt <- cpt.meanvar(
  dataset$Freq,
  method = "PELT",
  test.stat = "Poisson",
  penalty = "Hannan-Quinn",
  minseglen = 5
)

plot(meanvar.pelt, main = "Mean & Variance w/PELT", ylab = "Postings")
PlotGroundTruth(groundtruth.indexed)


# MeanVar SegNeigh
meanvar.segneigh <- cpt.meanvar(dataset$Freq,
                                method = "SegNeigh",
                                penalty = "Hannan-Quinn",
                                Q = 5)

plot(meanvar.segneigh, main = "Mean & Variance w/SegNeigh", ylab = "Postings")
PlotGroundTruth(groundtruth.indexed)


# MeanVar BinSeg
```

```
  meanvar.binseg <- cpt.meanvar(
    dataset$Freq,
    method = "BinSeg",
    test.stat = "Poisson",
    penalty = "Hannan-Quinn",
    Q = 5
  )

  plot(meanvar.binseg, main = "Mean & Variance w/BinSeg", ylab = "Postings")
  PlotGroundTruth(groundtruth.indexed)

}
```