GSoC: Periodic Time Changepoint Detection Easy Test

Haoting Tan

2024-04-01

Easy Test

Goal:

install the existing (Bayesian) periodic code from https://github.com/taylors2/PeriodCPT and run it on some binary data. Create fully reproducible code in Rmarkdown.

Install Necessary Packages

```
# Install devtools if not already installed
if (!requireNamespace("devtools", quietly = TRUE)) install.packages("devtools")

# set GITHUB_PAT to access github packages
if (!requireNamespace("PeriodCPT", quietly = TRUE)){
  Sys.setenv(GITHUB_PAT = "github_pat_11AP2TE5Y0G3nMiRDMf8LP_kYeGUKQIc5nFwL3A6Mi1TrixzC0D7NK571PrvQJpil
  # Install PeriodCPT from GitHub
  devtools::install_github("taylors2/PeriodCPT")
}

# Load the PeriodCPT package
library(PeriodCPT)
```

Generate Binary Data

```
set.seed(217) # For reproducibility
binary_data = ts( rbinom(90, size = 1, prob = rep(c(0.2, 0.8), each=6)), freq = 12)
```

Run PeriodCPT on the binary_data

```
# Bernoulli distribution
PeriodCPT(binary_data, distribution = "bern", niter = 1e3)
```

```
## Chain 1/1 (iteration):  |====================|
```

```
## Class 'pcpt' : Changepoint Object
##         ~~   : S4 class containing 18 slots with names
##                version data.set periodlength minseglen npcpts.max distribution nsegparam pcpt.prior pa
##
## Created on   : Mon Apr  1 18:57:08 2024
##
## summary(.)   :
## ----------
## Created Using changepoint version 1.2.1
## Distribution            :  bern
## Period length           :  12
## Minimum Segment Length  :  1
## Maximum no. of cpts     :  12
## Number of chains        :  1
## Number of periodic segs :  3
## Periodic cpt locations  :  6, 8, 11
## Seg. parameters at mode :
##             Seg1      Seg2 Seg3
## Param1 0.2181818 0.5714286    1
```

```
# poisson distribution
PeriodCPT(binary_data, distribution = "pois", niter = 1e3)
```

```
## Chain 1/1 (iteration):  |====================|
```

```
## Class 'pcpt' : Changepoint Object
##         ~~   : S4 class containing 18 slots with names
##                version data.set periodlength minseglen npcpts.max distribution nsegparam pcpt.prior pa
##
## Created on   : Mon Apr  1 18:57:16 2024
##
## summary(.)   :
## ----------
## Created Using changepoint version 1.2.1
## Distribution            :  pois
## Period length           :  12
## Minimum Segment Length  :  1
## Maximum no. of cpts     :  12
## Number of chains        :  1
## Number of periodic segs :  2
## Periodic cpt locations  :  6, 11
## Seg. parameters at mode :
##             Seg1      Seg2
## Param1 0.2142857 0.8055556
```

```
# normal distribution
PeriodCPT(binary_data, distribution = "norm", niter = 1e3)
```

```
## Chain 1/1 (iteration):  |====================|
```

```
## Class 'pcpt' : Changepoint Object
##         ~~   : S4 class containing 18 slots with names
```

```
##               version data.set periodlength minseglen npcpts.max distribution nsegparam pcpt.prior pa
##
## Created on   : Mon Apr  1 18:57:20 2024
##
## summary(.)   :
## ----------
## Created Using changepoint version 1.2.1
## Distribution            :  norm
## Period length           :  12
## Minimum Segment Length  :  1
## Maximum no. of cpts     :  12
## Number of chains        :  1
## Number of periodic segs :  2
## Periodic cpt locations  :  8, 11
## Seg. parameters at mode :
##             Seg1      Seg2
## Param1 0.2857143 0.9545455
## Param2 0.2200772 0.1136364
```

Summary

- The summaries you've provided show the results of running the PeriodCPT algorithm on the same binary data with three different assumed data distributions: Bernoulli (bern), Poisson (pois), and Normal (norm). These summaries give insight into how the choice of distribution affects the detected changepoints, the estimated segment parameters, and the interpretation of the underlying data structure.