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
Sys.setenv(GITHUB_PAT = "github_pat_11AP2TE5Y0G3nMiRDMf8LP_kYeGUKQIc5nFwL3A6Mi1TrixzC0D7NK571PrvQJpilrF

# Install PeriodCPT from GitHub
devtools::install_github("taylors2/PeriodCPT")
```

```
## Using github PAT from envvar GITHUB_PAT. Use `gitcreds::gitcreds_set()` and unset GITHUB_PAT in .Renv
```

```
## Skipping install of 'PeriodCPT' from a github remote, the SHA1 (fd8c5f1f) has not changed since last
##   Use `force = TRUE` to force installation
```

```
# 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

```
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   : Sun Mar 31 21:23:11 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
```

Summary

- The summary indicates that the binary data exhibits periodic behavior with significant variations in patterns across different segments within the specified period length. Specifically, it identifies three distinct segments within each period, where the likelihood of observing a 1 changes markedly. This could imply that the process generating the data has different states or behaviors that switch at specific points in time, and these changes are periodic, repeating every 12 observations.