STOMP motif discovery: Matrix Profile methods compute, for each subsequence, the distance to its nearest neighbor subsequence. STOMP is an efficient exact algorithm for long series. Harbinger provides Matrix Profile via tsmp and wraps it in hmo\_mp().

Objectives: This Rmd demonstrates motif discovery using Matrix Profile with the STOMP algorithm via hmo\_mp("stomp", ...). It identifies repeated subsequences (motifs) efficiently in long series. Steps: load packages/data, visualize, configure subsequence length and number of motifs, fit, detect, evaluate, and plot.

# Install Harbinger (only once, if needed)  
#install.packages("harbinger")

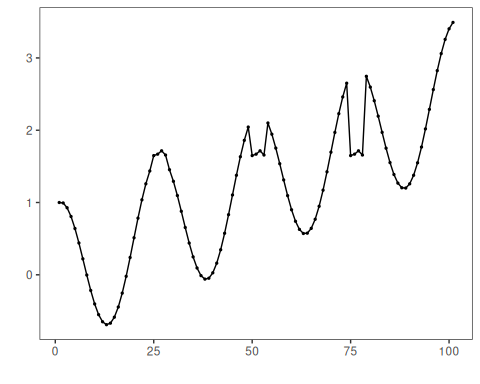
# Load required packages  
library(daltoolbox)  
library(harbinger)

# Load example datasets bundled with harbinger  
data(examples\_motifs)

# Select a simple example time series  
dataset <- examples\_motifs$simple  
head(dataset)

## serie event  
## 1 1.0000000 FALSE  
## 2 0.9939124 FALSE  
## 3 0.9275826 FALSE  
## 4 0.8066889 FALSE  
## 5 0.6403023 FALSE  
## 6 0.4403224 FALSE

# Plot the time series  
har\_plot(harbinger(), dataset$serie)



# Define Matrix Profile (STOMP) motif model  
# - second arg: subsequence length (window)  
# - third arg: number of motifs to retrieve  
 model <- hmo\_mp("stomp", 4, 3)

# Fit the model  
 model <- fit(model, dataset$serie)

# Detect motifs  
 detection <- detect(model, dataset$serie)

## Finished in 0.01 secs

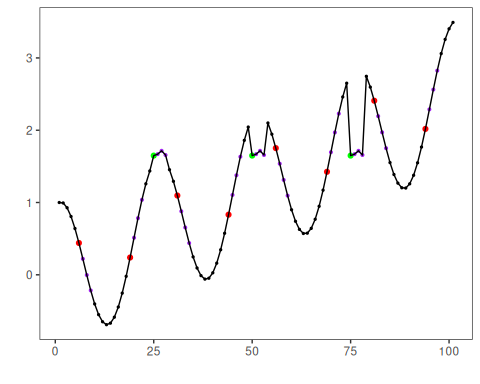
# Show only timestamps flagged as events  
 print(detection |> dplyr::filter(event==TRUE))

## idx event type seq seqlen  
## 1 6 TRUE motif 3 4  
## 2 19 TRUE motif 2 4  
## 3 25 TRUE motif 1 4  
## 4 31 TRUE motif 3 4  
## 5 44 TRUE motif 2 4  
## 6 50 TRUE motif 1 4  
## 7 56 TRUE motif 3 4  
## 8 69 TRUE motif 2 4  
## 9 75 TRUE motif 1 4  
## 10 81 TRUE motif 3 4  
## 11 94 TRUE motif 2 4

# Evaluate detections against ground-truth labels  
 evaluation <- evaluate(model, detection$event, dataset$event)  
 print(evaluation$confMatrix)

## event   
## detection TRUE FALSE  
## TRUE 3 8   
## FALSE 0 90

# Plot detections over the series  
 har\_plot(model, dataset$serie, detection, dataset$event)



References - Yeh, C.-C. M., et al. (2016). Matrix Profile I/II: All-pairs similarity joins and scalable time series motif/discord discovery. IEEE ICDM. - Tavenard, R., et al. (2020). tsmp: The Matrix Profile in R. The R Journal. <doi:10.32614/RJ-2020-021>