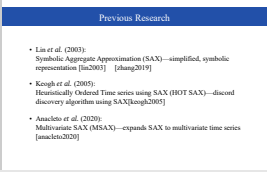
Previous Research

- Lin *et al.* (2003):
  Symbolic Aggregate Approximation (SAX)—simplified, symbolic representation [lin2003] [zhang2019]
- Keogh *et al.* (2005):
  Heuristically Ordered Time series using SAX (HOT SAX)—discord discovery algorithm using SAX[keogh2005]
- Anacleto *et al.* (2020):
  Multivariate SAX (MSAX)—expands SAX to multivariate time series [anacleto2020]



- ecg as letters that mean same thing as original
- guaranteed to behave like the original data
- works on univariate time series
- has been used on ECGs
- uses sax representation to make the finding of discords easier
- can use MSAX just as well
- takes the correlation between ecg leads into account
- cov mat: covariance between each lead and variance on diag

This Work's Novel Contributions

- application of MSAX to ECG discord discovery
- the HOT MSAX algorithm
- the expansion of HOT SAX to multivariate time series through HOT MSAX

SAX and MSAX – Overview

SAX	MSAX
Application	
univariate time series e.g. a single ECG lead	multivariate time series e.g. multiple ECG leads
Steps	
(1) univariate z-normalization (2) PAA dimension reduction (3) SAX discretization	(1) multivariate z-normalization (2) PAA dimension reduction (3) SAX discretization

SAX and MSAX – Step (2)

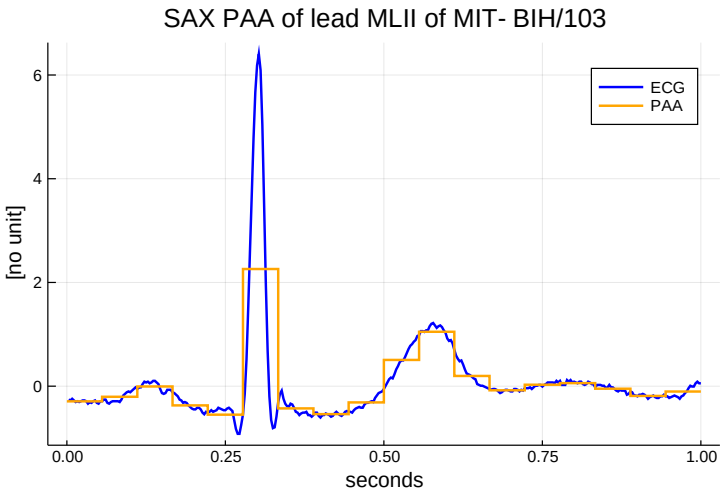


Figure 1: ECG with PAA (MIT-BIH/103, $w = 18$, $T = 360$)

SAX and MSAX – Step (3)

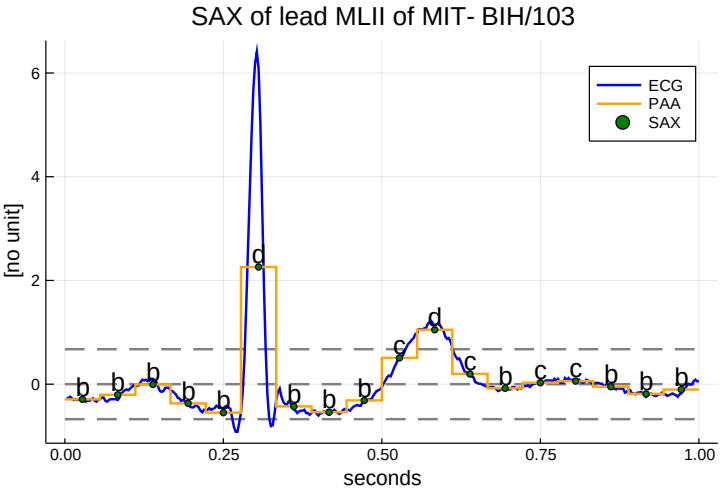


Figure 2: ECG with SAX (MIT-BIH/103, $w = 18$, $T = 360$)

SAX and MSAX – Distance Measure

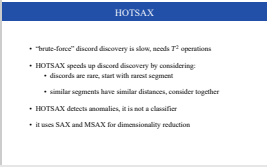
- two SAX or MSAX segments can be compared (“distance”)
- basically the sum of distances between symbols
-

HOT SAX and HOT MSAX – Overview

- HOT SAX: find discords in SAX-represented time series
- speeds up the “brute force” approach
- classifies time series segments into “discord” and “non-discord”
- HOT MSAX: uses MSAX instead of SAX
- HOT MSAX can work with multivariate time series

HOTSAX

- "brute-force" discord discovery is slow, needs T^2 operations
- HOTSAX speeds up discord discovery by considering:
  - discords are rare, start with rarest segment
  - similar segments have similar distances, consider together
- HOTSAX detects anomalies, it is not a classifier
- it uses SAX and MSAX for dimensionality reduction



- this is the basic idea that can speed up the process
- it is not guaranteed to do so, but it does not decrease efficiency
- this speeds up the process even more as we have fewer elements
- because of lower bounding, it still gives accurate results