Symbolic representations for time series

PhD defense

Sylvain W. Combettes

Supervisors: Laurent Oudre and Charles Truong

January 8th, 2024

1 – Introduction

Context
Centre Borelli

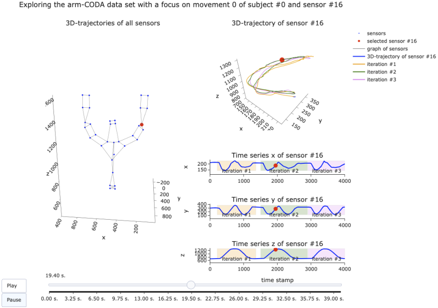
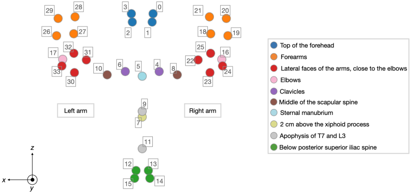


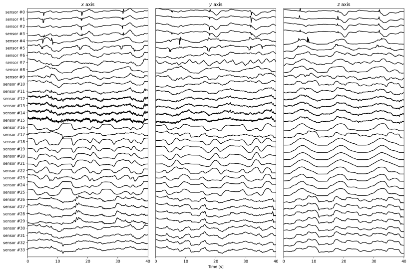
Figure: armCODA data set.

- Neuroscience projects: often combining mathematicians with medical doctors and clinicians.
- Analysis of human behavior
  1. **Longitudinal follow-up**: studying the evolution of a subject over time.
  2. **Inter-individual comparison**: comparing two cohorts of subjects.
- Creation of data sets of physiological signals from protocols
  - armCODA data set [1]: study of arm movements
  - gait data set [9]: study of human locomotion

Context

Use case #1: armCODA data set [1]



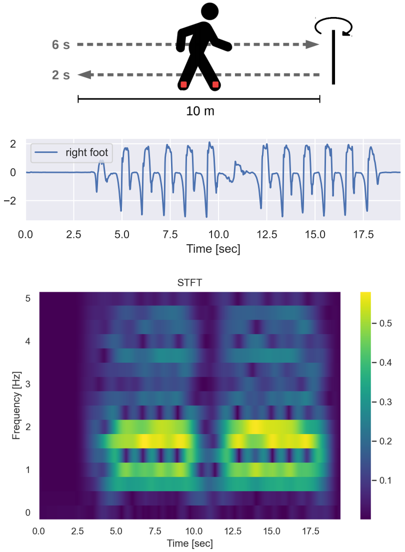


- Goal: study of upper-limb movements during rehabilitation after injury
- 34 CODA sensors (Cartesian Optoelectronic Dynamic Anthropometer), recording the 3D position, placed on the upper limb of 16 patients
- Protocol: patients performing 15 movements
  - raising their arms
  - combing their hair
  - ...

↪ 240 multivariate signals with **102 dimensions**

Context

Use case #2: gait data set [9]



- Goal: study of human locomotion for early detection of fall risk
- Sensors: angular velocity recorded on the left and right feet using a pair of sensors.
- Protocol: standing, walking, turning around, walking back, and standing.
- Preprocessing: norms of the STFT (Short Time Fourier Transform) of each foot recording (univariate signal)

➥ 442 multivariate signals with **16 dimensions**

Scientific questions and challenges

- Scientific questions
  1. How to **represent** physiological signals with a complex structure?
  2. How can we define a **distance** between them?
- Challenges
  - temporal information: retain the chronology of actions
  - noise
  - multivariate/multimodal: many dimensions (e.g. 102), possibly correlated
  - non-stationary: statistical properties of the signals change over time
  - computational cost
  - interpretability for clinicians

Our goals and our approach

- Our goals when representing and comparing complex physiological signals
  - Adapt to the phenomena of interest.
  - Perform the comparison at the level of "actions".
  - Be fast to compute (almost interactive).
  - Allow longitudinal follow-up and inter-individual comparison.
- Our approach
  1. Symbolization: transforming a real-valued series into a shorter discrete-valued series.

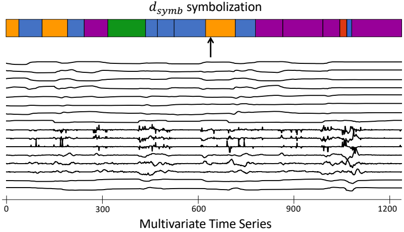


Figure: Example of symbolization.

  2. Applying a distance measure on the resulting strings.

2 – Background and related work

Background and related work

- In the manuscript, we have conducted two literature reviews:
  - Chapter II: Symbolic representations for time series.
    Covers more than 60 symbolization methods.
  - Chapter III: Distance measures on time series, strings, and symbolic sequences.
    - A *time series* is a series of real values indexed in time order.
    - A *string* is a series of discrete values indexed in time order, the discrete values being non-ordered and taken from a fixed alphabet of characters.
    - A *symbolic sequence* is a discrete sequence resulting from the transformation of a time series using a symbolization process.

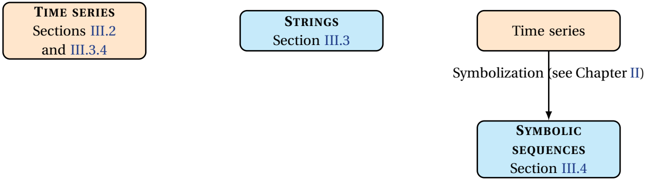


Figure: Overview of distance types reviewed in the manuscript.

Symbolic representation of time series
Framework

Symbolization of a time series:

1. **Segmentation**: a real-valued signal $y = (y_1, \dots, y_n)$ of length n is split into w segments ($w < n$)
2. **Feature extraction**: features of interest are extracted for each segment
3. **Quantization** (of the real-valued extracted features): each segment is mapped to a discrete value taken from a set $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \dots\}$ of A symbols

Figure: Main steps for the symbolization of a time series.

Notations and vocabulary:

- word length (number of segments): w
- alphabet size (number of symbols): A
- alphabet (a.k.a dictionary): $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \dots\}$ or $\{0, 1, 2, \dots\}$

Symbolic representation of time series

A popular method: Symbolic Aggregate approXimation (SAX) [6]

1. Segmentation: uniform, with the word length w
2. Feature extraction: mean
3. Quantization: Gaussian bins, with alphabet size A

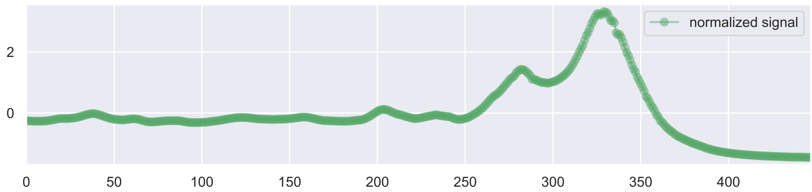


Figure: Example of SAX [6] representation of a univariate signal, with $w = 4$ and $A = 4$.

Symbolic representation of time series

A popular method: Symbolic Aggregate approXimation (SAX) [6]

1. Segmentation: uniform, with the word length w
2. Feature extraction: mean
3. Quantization: Gaussian bins, with alphabet size A

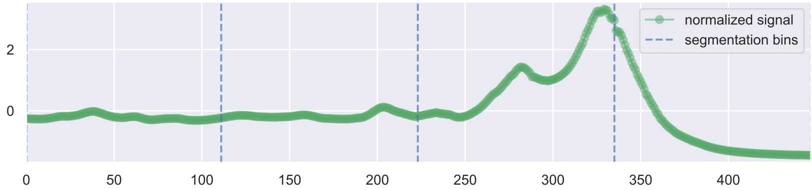


Figure: Example of SAX [6] representation of a univariate signal, with $w = 4$ and $A = 4$.

Symbolic representation of time series

A popular method: Symbolic Aggregate approXimation (SAX) [6]

1. Segmentation: uniform, with the word length w
2. Feature extraction: mean
3. Quantization: Gaussian bins, with alphabet size A

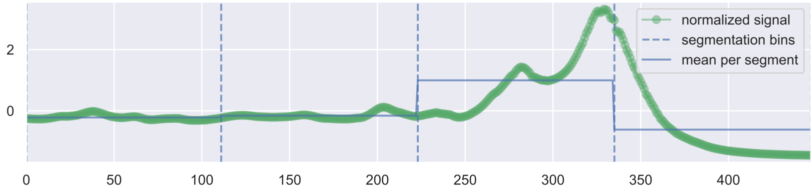


Figure: Example of SAX [6] representation of a univariate signal, with $w = 4$ and $A = 4$.

Symbolic representation of time series

A popular method: Symbolic Aggregate approXimation (SAX) [6]

1. Segmentation: uniform, with the word length w
2. Feature extraction: mean
3. Quantization: Gaussian bins, with alphabet size A

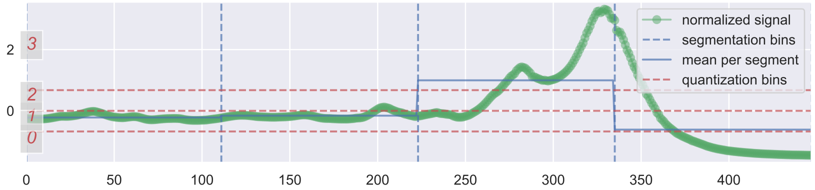


Figure: Example of SAX [6] representation of a univariate signal, with $w = 4$ and $A = 4$.

Symbolic representation of time series

A popular method: Symbolic Aggregate approXimation (SAX) [6]

1. Segmentation: uniform, with the word length w
2. Feature extraction: mean
3. Quantization: Gaussian bins, with alphabet size A

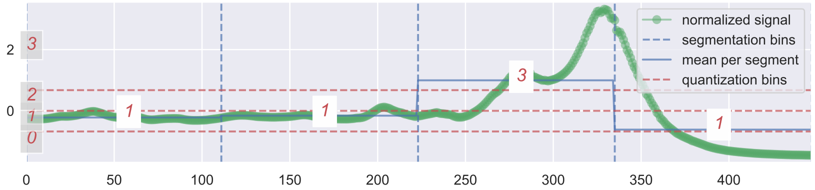


Figure: Example of SAX [6] representation of a univariate signal, with $w = 4$ and $A = 4$.

Symbolic representation of time series

A popular method: Symbolic Aggregate approXimation (SAX) [6]

1. Segmentation: uniform, with the word length w
2. Feature extraction: mean
3. Quantization: Gaussian bins, with alphabet size A

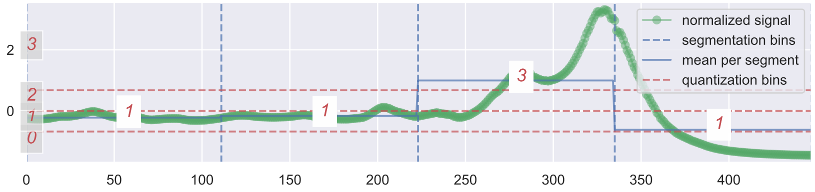


Figure: Example of SAX [6] representation of a univariate signal, with $w = 4$ and $A = 4$.

- Applications: clustering, classification, query by content, anomaly detection, motif discovery, and visualization.

Symbolic representation of time series

Some popular methods

- Variants of SAX in the literature: modify one or more steps.

Table: Summary of some popular symbolic representations.

Method	Segmentation	Feature extraction	Quantization
Symbolic Aggregate approXimation (**SAX**) [6]	uniform	mean	Gaussian bins
1d-SAX [7]	uniform	mean, slope	Gaussian bins
Symbolic Fourier Approximation (**SFA**) [8]	∅	Fourier coefficients	quantiles
Adaptive Brownian Bridge-based Aggregation (**ABBA**) [3]	piecewise linear approximation	increment, length	clustering

Distance measures on series

On time series

- L_p distance between $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$

$$L_p(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}$$

- DTW (Dynamic Time Warping) and variants: robust to time-shifts

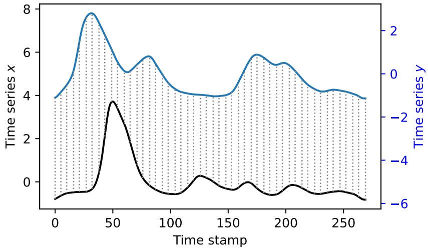


Figure: Euclidean distance: one-to-one alignment. Sample x_i is associated with sample y_i.

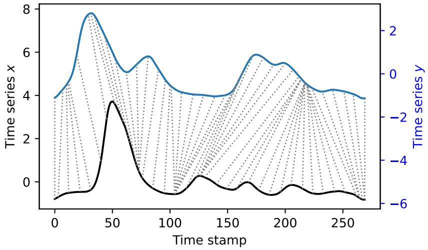


Figure: DTW distance: one-to-many alignment. Sample x_{i_k} is associated with sample y_{j_k}.

Distance measures on series
On strings

- Edit distance on strings: minimal cost of a sequence of operations that transform a string into another.
- Allowed simple operations:
  - Insertion: abc $\longrightarrow$ abcd
  - Deletion: abc $\longrightarrow$ ac
  - Substitution: abc $\longrightarrow$ adc
  - Transposition: ab $\longrightarrow$ ba
  - Duplication: abc $\longrightarrow$ abbc
  - Contraction: abbc $\longrightarrow$ abc
- Cost of a simple operation: depends on
  - operation type
  - characters involved
- Total cost: sum of the costs of the simple operations.

Distance measures on series

On strings

Table: Summary of edit distances on strings of lengths *m* and *n*.
†Depends on how the operation costs are set.

Distance name	Insertion	Deletion	Substitution	Transposition	Duplication	Contraction	Time complexity
	Allowed edit operations						Property
LCSS [Hir77]	✔	✔	✘	✘	✘	✘	$\mathcal{O}(mn)$
Hamming [SM83]	✘	✘	✔	✘	✘	✘	$\mathcal{O}(m)$
Simple Levenshtein distance [Lev+66]	✔	✔	✔	✘	✘	✘	$\mathcal{O}(mn)$
General Levenshtein distance [Lev+66]	✔	✔	✔	✘	✘	✘	$\mathcal{O}(mn)$
Damerau-Levenshtein	✔	✔	✔	✔	✘	✘	$\mathcal{O}(mn)$
Edit Distance with Duplications and Contractions (EDDC) [BR02; Pin+13]	✔	✔	✔	✘	✔	✔	$\mathcal{O}\left(\lvert\mathscr{A}\rvert m^3\right)$

Distance measures on series
On symbolic sequences

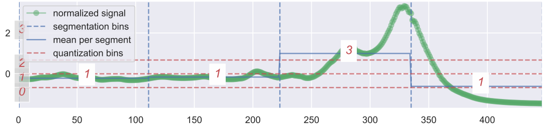


Figure: Example of SAX representation with $w = 4$ and $A = 4$.

- MINDIST distance (from SAX) between symbolic sequences $\hat{x}$ and $\hat{y}$:

$$D_{\text{MINDIST}}(\hat{x}, \hat{y}) = \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^{w} \left(\text{dist}(\hat{x}_i, \hat{y}_i)\right)^2}$$

where the dist function is based on a look-up table:

Table: Example of look-up table for MINDIST with $A = 4$ for the quantization bins β_i.

	a	b	c	d
a	0	0	$\beta_2 - \beta_1$	$\beta_3 - \beta_1$
b	0	0	0	$\beta_3 - \beta_2$
c	$\beta_2 - \beta_1$	0	0	0
d	$\beta_3 - \beta_1$	$\beta_3 - \beta_2$	0	0

3 – ASTRIDE: for univariate time series

Limitations of existing symbolization methods

The need for adaptive segmentation and quantization steps

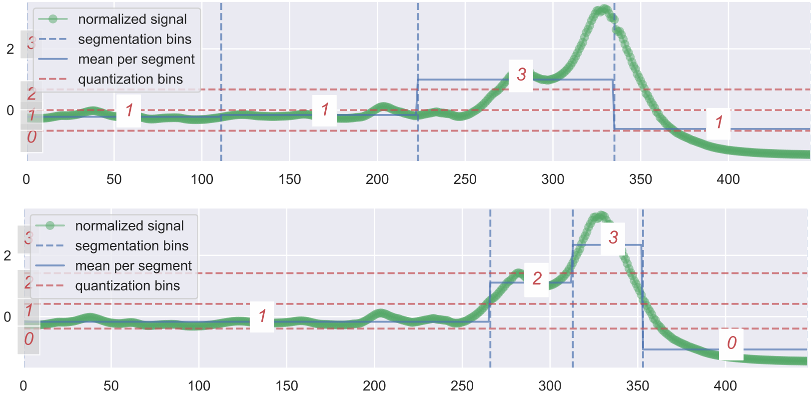


Figure: Example of SAX (top) and ASTRIDE (bottom) representations of a signal with $n = 448$, $w = 4$, and $A = 4$.

- ✘ Uniform segmentation can not detect salient events such as peaks.
- ✘ Fixed (Gaussian) bins are not data-driven.

Limitations of existing symbolization methods

The need for a distance measure on symbolic sequences

Table: Summary of some popular symbic representations.

Method	Feature extraction	Adaptive segmentation?	Adaptive quantization?	Distance measure?
SAX [6]	mean	✘	✘	✔
1d-SAX [7]	mean, slope	✘	✘	✔
SFA [8]	$\varnothing$	$\varnothing$	✔	✘
ABBA [3]	increment, length	✔	✔	✘
ASTRIDE	mean	✔	✔	✔

- Many symbolic representations do not hold a distance measure.
- MINDIST from SAX...
  - considers adjacent symbols to be equal
  - is based on the fixed Gaussian assumption
  - is restricted to equal-length symbolic sequences

Limitations of existing symbolization methods

The need for a shared dictionary of symbols across the signals of a data set

- Task: reconstruction.
  - Symbolization: compression
    - of N time series with n samples each, each sample being encoded on n_{bits} bits
    - into N discrete-values series with w samples each, each sample being encoded on $\log_2(A)$ bits.
  - Reconstruction: decompression.

Table: Memory usage (in bits) to reconstruct N symbolic sequences.

Method	N symbolic sequences	Dictionaries of A symbols (for all N signals)
Raw time series	$Nn n_{\text{bits}}$	
SAX	$Nw \log_2(A)$	$n_{\text{bits}} A$
ABBA	$Nw \log_2(A)$	$2 n_{\text{bits}} N A$

Table: Meat data set (UCR archive [2]) with $N = 120$, $n = 448$, $w = 10$, $A = 9$, and $n_{\text{bits}} = 64$ bits.

Method	Raw time series	SAX	ABBA
Nb of bits	3,440,640	4,380	142,044

➥ ABBA requires much more memory usage than SAX (e.g. 32 times more) because it is adaptive and its dictionary of symbols is not shared across signals.

The ASTRIDE method

Adaptive segmentation step

Stacking: from N univariate signals to 1 multivariate signal of dimension N.

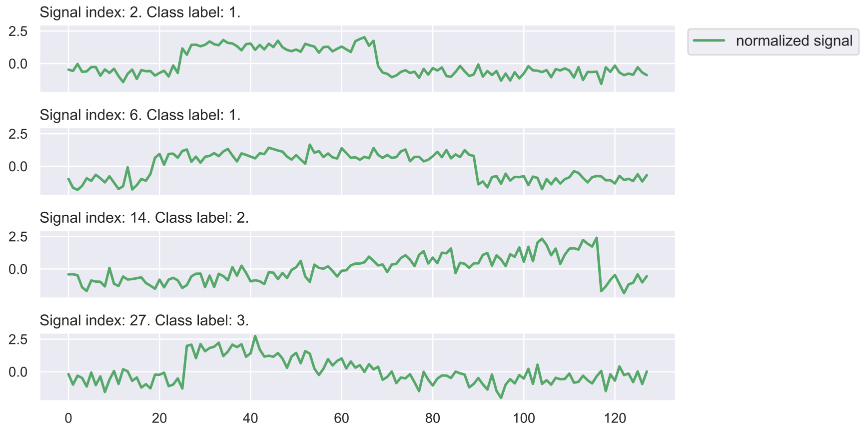


Figure: Stacking univariate signals with $n = 128$.

The ASTRIDE method

Adaptive segmentation step

- Change-point detection: finds the $w-1$ unknown instants $t_1^* < t_2^* < \ldots < t_{w-1}^*$ where the mean of $y = (y_1, \ldots, y_n)$ of dimension N changes abruptly

$$(\hat{t}_1, \ldots, \hat{t}_{w-1}) = \underset{(t_1, \ldots, t_{w-1})}{\arg\min} \sum_{k=0}^{w+1} \sum_{t=t_k}^{t_{k+1}-1} \left\| y_t - \bar{y}_{t_k:t_{k+1}} \right\|^2$$

where $\bar{y}_{t_k:t_{k+1}}$ is the empirical mean of $\{y_{t_k}, \ldots, y_{t_{k+1}-1}\}$.
- w is the user-chosen number of segments.
- The formulation seeks to reduce the error between the original signal and the best piecewise constant approximation.
- Solved using dynamic programming with a time complexity of $\mathcal{O}\left(Nwn^2\right)$.

The ASTRIDE method

Adaptive segmentation step

Stacking: from N univariate signals to 1 multivariate signal of dimension N, so the change-points are shared thus memory-efficient.

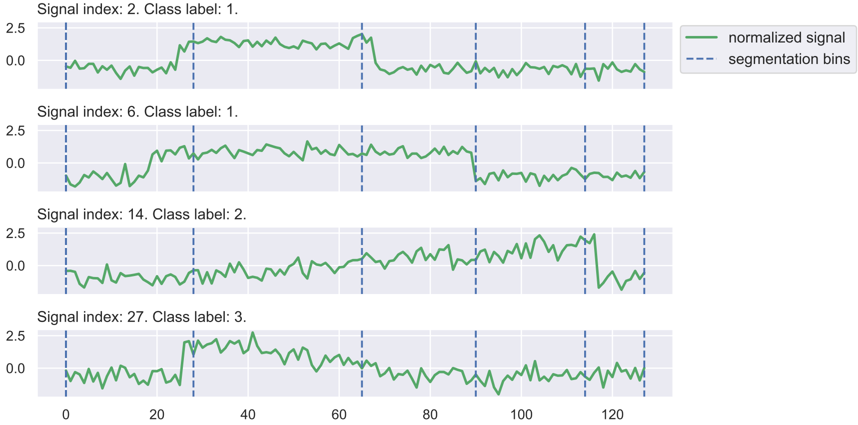


Figure: Multivariate change-point detection on (univariate) signals with $n = 128$ and $w = 5$.

The ASTRIDE method

Adaptive quantization step

- Quantization bins: empirical quantiles of the means of all segments.
- Remarks
  - The segmentation corresponds to mean-shifts, so we represent each segment by its mean value.
  - By design, all symbols are equiprobable.
  - ➥ Shared dictionary of symbols: all steps are learned on a whole data set, thus ASTRIDE is memory-efficient.

The ASTRIDE method

The D-GED (Dynamic General Edit Distance) distance measure

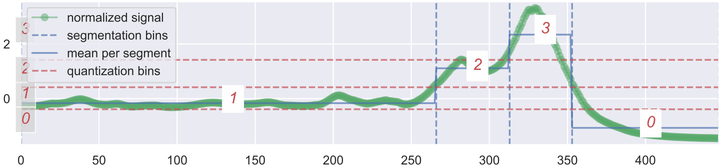


Figure: Example of ASTRIDE representation of a signal with $n = 448$, $w = 4$, and $A = 4$.

1. Preprocessing.
   - Including the segment length information: replicating each symbol proportionally to its segment length.
     Example: 1230, with lengths 8, 2, 2, and 4 becomes 1111111122330000.
   - Shortening: dividing each length by the minimum length.
     Example: 1111111122330000 becomes 11112300.
2. Applying the general edit distance with custom costs.
   - Substitution: Euclidean distance between the average mean values of the symbols.
   - Insertion: max of substitution costs.
   - Deletion: max of substitution costs.

The ASTRIDE method

Reconstruction of the ASTRIDE symbolic sequences

1. Each symbol is replicated by its true length.
2. Each symbol is replaced by its corresponding average of extracted mean features.

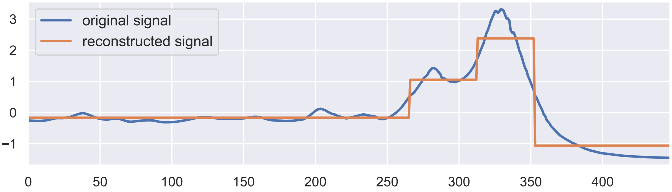


Figure: Example: reconstruction by ASTRIDE of a symbolic sequence with $w = 4$ and $A = 4$.

- Memory cost: $\boxed{Nw \log_2(A) + (w + A) n_{\text{bits}}}$ bits.

Table: Nb of bits to reconstruct a data set with $N = 120$, $w = 10$, $A = 9$, and $n_{\text{bits}} = 64$.

SAX	ABBA	ASTRIDE
$4,380$	$142,044$	$5,020$

➥ ABBA takes 28 times more bits than ASTRIDE.

The ASTRIDE method
FASTRIDE

FASTRIDE (Fast ASTRIDE): accelerated variant of ASTRIDE.

Table: Comparing ASTRIDE and FASTRIDE.

Method	ASTRIDE	FASTRIDE
Segmentation	adaptive	**uniform**
Quantization	quantiles	quantiles
Distance	D-GED on replicated symbols	D-GED on **unreplicated** symbols

Experimental results

Task	Classification	Reconstruction
Score	1-nearest neighbor classification accuracy	reconstruction error (Euclidean and DTW)
Benchmark	SAX, 1d-SAX, ASTRIDE, FASTRIDE	SAX, 1d-SAX, SFA, ABBA, ASTRIDE, FASTRIDE
Data sets	univariate and equal-size times series from the UCR Times Series Classification Archive [2]	
Nb of data sets	86	60

Table: Experimental setup

- Python implementation: https://github.com/sylvaincom/astride
- ➥ Results: ASTRIDE and FASTRIDE are the best for classification, and second best for reconstruction (after SFA).

Experimental results

Classification task

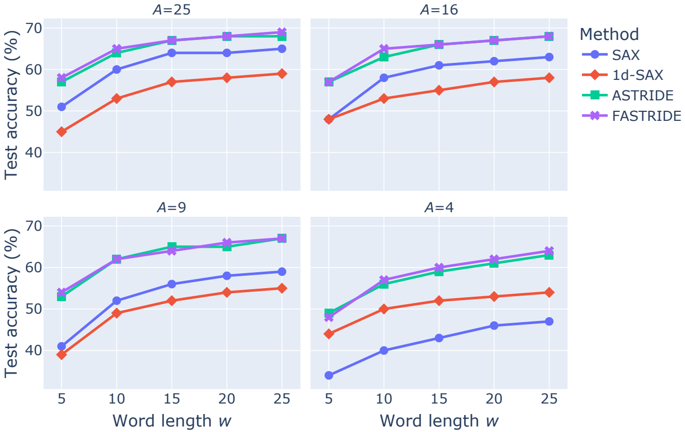


Figure: Classification benchmark averaged on 86 data sets from the UCR archive.

- ASTRIDE and FASTRIDE (quite similar) perform better than both SAX and 1d-SAX, and are quite robust to low values of w.

Experimental results

Reconstruction task

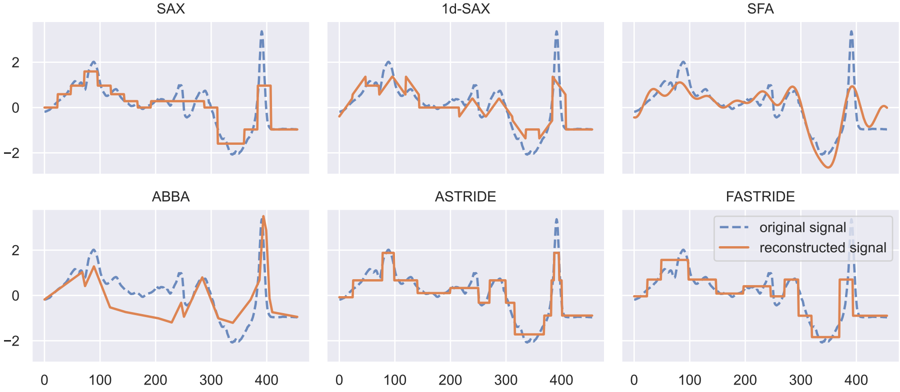


Figure: Example of reconstruction of a signal with $n = 470$, $A = 9$ and $w = 19$.

- ASTRIDE seems to perform better on this particular signal: SFA does not account well for peaks and ABBA has quantized segment lengths.

Experimental results
Reconstruction task

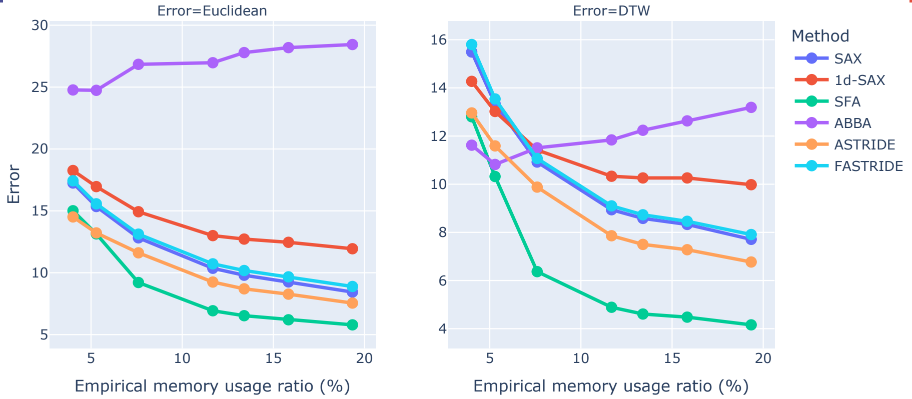


Figure: Benchmarking the reconstruction error, averaged on around 60 data sets from the UCR archive, with $A = 9$, with regards to the empirical memory usage ratio being w/n.

- ASTRIDE performs 2nd best behind SFA (and better than FASTRIDE).
- For very low memory usage ratios, ASTRIDE is competitive with SFA.

Experimental results

Computational complexity

Table: Processing times (in sec) of the symbolization, 1-NN classification, and reconstruction on the ECG200 data set composed of 100 training signals and 100 test signals of length $n = 96$, with $w = 10$ and $A = 9$.

Method	symbolization	1-NN classification
SAX	0.27	0.08
SAX (`tslearn`)	0.02	0.11
1d-SAX (`tslearn`)	0.42	0.21
ASTRIDE	0.30	0.17
FASTRIDE	0.26	0.07

- ➥ The adaptive segmentation step is quite fast (ASTRIDE vs FASTRIDE).
- ➥ The classification of FASTRIDE is faster than ASTRIDE due to the unreplicated symbolic sequences.

4 – d_symb: for multivariate time series

Limitations of existing approaches

- Distance measures on multivariate time series $\rightarrow$ extensions of distances in univariate time series with 2 strategies:
  - Independent strategy: summing the univariate distances from all dimensions
  - Dependent strategy: for example, in DTW, a multivariate series is considered as a single series where each timestamp is a multidimensional point
  - ✘ Computational cost, interpretability.
- Symbolic representations for multivariate time series $\rightarrow$ rare
  - Dimensionality reduction: apply PCA then symbolize the univariate reduced signal
  - Independent strategy: symbolize each dimension independently, then
    - concatenates them into a single long string
    - uses a multivariate Gaussian distribution with a total alphabet of size A^d, with d the dimension
    - ✘ do not scale well with the dimension d, interpretability of (large) alphabets
  - Dependent strategy: multivariate version of the mean per segment of SAX: real value that corresponds to the average of the L_2-norms of each multidimensional sample

The d_symb symbolization and distance measure

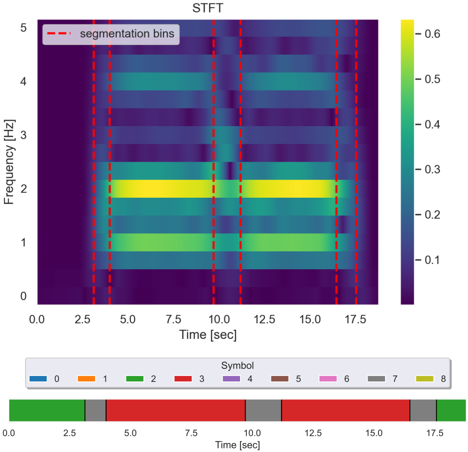


Figure: Multivariate signal (spectrogram) and its d_{symb} symbolic sequence.

Steps of d_{symb}

1. Segmentation: change-point detection (on the mean).
2. Quantization: K-means clustering (of the mean vectors per segment), with $K = A$.
3. Distance: general edit distance between the resulting symbolic signals.

The d_symb symbolization and distance measure

Segmentation

- Change-point detection: finding the w^* unknown instants $t_1^* < t_2^* < \ldots < t_{w^*}^*$ where the mean of signal $x = (x_1, \ldots, x_n)$ change abruptly:

$$\left(\hat{w}, \hat{t}_1, \ldots, \hat{t}_{\hat{w}}\right) = \underset{(w, t_1, \ldots, t_w)}{\arg\min} \sum_{k=0}^{w+1} \sum_{t=t_k}^{t_{k+1}-1} \left\| x_t - \bar{x}_{t_k:t_{k+1}} \right\|^2 + \lambda w$$

  where $\bar{x}_{t_k:t_{k+1}}$ is the empirical mean of $\left\{x_{t_k}, \ldots, x_{t_{k+1}-1}\right\}$ and $\lambda > 0$ is a penalization parameter.

- Compromise between the reconstruction error and the number of change-points.
- When λ is small, many change-points are detected.
  For calibration purposes, we often use $\lambda = \ln(n)$ [10].
- Solved using the Pruned Exact Linear Time (PELT) algorithm [5], which is shown to have $\mathcal{O}(n)$ complexity (under some assumptions).

The d_symb symbolization and distance measure

Distance measure

1. Preprocessing as in ASTRIDE.
   - Replicating each symbol proportionally to its segment length.
   - Shortening.
2. Applying the general edit distance with custom costs.
   - Substitution: Euclidean distance between the cluster centers of the symbols.
   - Insertion: max of substitution costs.
   - Deletion: max of substitution costs.

Experimental results

Application of d_{symb} to 3 real-world data sets of multivariate physiological signals

Data set	Data set description	N	d	Experimental task
Human locomotion [9]	standing, **walking**, turning around	442	16	interpretation
armCODA [1]	**arm elevation**	240	102	interpretation
JIG SAWS [4]	**surgical tasks** performed by 8 surgeons using robotic arms and grippers, with a focus on 2 gestures: knot tying and needle passing	79	76	clustering, interpretation

Table: Experimental setup

➥ Results: d_{symb} is fast to compute and is interpretable.

Experimental results

Human locomotion data set

Figure: Color bars for 60 recordings, with $\lambda = \ln(n)$ and $A = 9$.

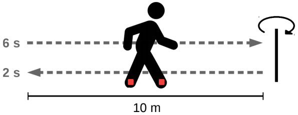


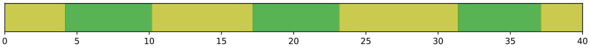
Figure: Protocol

Interpretation of the d_{symb} symbolization:

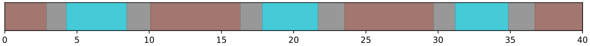
- The general structure is coherent with the protocol.
- Change-point detection finds stationary segments.
- Each symbol can be associated with a type of behavior.

Experimental results

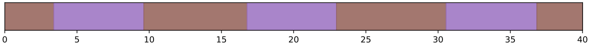
armCODA data set



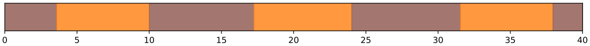
(a) seated, bilateral



(b) standing, bilateral



(c) standing, unilateral right



(d) standing, unilateral left

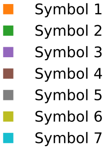
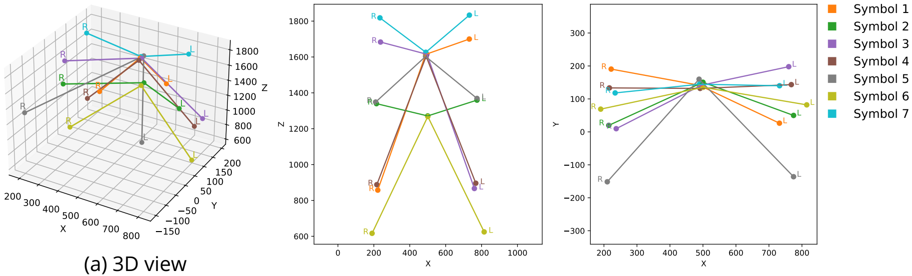


Figure: d_{symb} with $A = 7$. Same subject with 4 movements in sagittal plane elevation.

- We detect the 3 iterations of the protocol.
- Symbol 4: resting while standing. Symbol 6: resting while seating.
- Each movement has its own symbol.

Experimental results
armCODA data set



(a) 3D view (b) Front view (c) Top view

Figure: Positions (x, y, z) (in cm and in the laboratory frame) of the head, left forearm (L), and right forearm (R) for each symbol centroid.

- Each cluster center is an average of body positions.
- (Front view) Symbol 4: resting while standing. Symbol 6: resting while seating.
- (Front view) Symbol 7: bilateral arm elevation. Symbol 1: left arm elevation.

The d_symb playground

Demo time: application of d_symb to the JIG SAWS data set

- Streamlit app
  https://dsymb-playground.streamlit.app
- Python implementation
  https://github.com/boniolp/dsymb-playground

5 – Conclusion

Recap

- ASTRIDE: for a data set of univariate time series
  - ➥ Performs very well in classification and reconstruction, while being memory-efficient.

S. W. Combettes, C. Truong, and L. Oudre. "SAX-DD : une nouvelle représentation symbolique pour séries temporelles." Published in *Proceedings of the Groupe de Recherche et d'Etudes en Traitement du Signal et des Images (GRETSI)*, Nancy, France, September 2022.

S. W. Combettes, C. Truong, and L. Oudre. "ASTRIDE: Adaptive Symbolization for Time Series Databases." Submitted to *Data Mining and Knowledge Discovery (DAMI)* in February 2023.

- d_{symb}: for a data set of multivariate time series; showcased with the d_{symb} playground
  - ➥ Can deal with multivariate non-stationary physiological signals thanks to a change-point detection procedure.
  - ➥ Interpretable.
  - ➥ Much faster than DTW.

S. W. Combettes, C. Truong, and L. Oudre. "An Interpretable Distance Measure for Multivariate Non-Stationary Physiological Signals." To be published in *Proceedings of the International Conference on Data Mining Workshops (ICDMW)*, Shanghai, China, December 2023.

S. W. Combettes, P. Boniol, C. Truong, and L. Oudre. "d_{symb} playground: an interactive tool to explore large multivariate time series datasets." To be published in *Proceedings of the International Conference on Data Engineering (ICDE) – Demonstration track*, Utrecht, Netherlands, May 2024.

Perspectives

- Apply ASTRIDE or d_{symb} to more tasks
  - Intermediate step in classifiers
  - Analyzed by methods in bioinformatics
  - Markov chains
- Extension to even more complex physiological signals
  - Multi-resolution
  - Correlation between dimensions
- Investigate the distance
  - Links between edit distances and DTW?
  - Lower-bound?
- Multimodal aspect

Thank you for your attention.

References I

S. W. Combettes, P. Boniol, A. Mazarguil, D. Wang, D. Vaquero-Ramos, M. Chauveau, L. Oudre, N. Vayatis, P.-P. Vidal, A. Roren, and M.-M. Lefèvre-Colau.
Arm-CODA: A Dataset of Upper-limb Human Movement during Routine Examination.
Image Processing On Line (preprint), 2023.
https://www.ipol.im/pub/pre/494/.

H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh.
The ucr time series archive.
IEEE/CAA Journal of Automatica Sinica, 6(6):1293–1305, 2019.

S. Elsworth and S. Güttel.
Abba: adaptive brownian bridge-based symbolic aggregation of time series.
Data Min Knowl Disc, 34:1175–1200, 2020.

References II

Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh, et al.
The jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling.
In *Modeling and Monitoring of Computer Assisted Interventions (M2CAI) – MICCAI Workshop*, 2014.

R. Killick, P. Fearnhead, and I. A. Eckley.
Optimal detection of changepoints with a linear computational cost.
Journal of the American Statistical Association, 107(500):1590–1598, 2012.

J. Lin, E. Keogh, L. Wei, and S. Lonardi.
Experiencing sax: a novel symbolic representation of time series.
Data Min Knowl Disc, 15:107–144, 2007.

S. Malinowski, T. Guyet, R. Quiniou, and R. Tavenard.
1d-sax: A novel symbolic representation for time series.
In *Advances in Intelligent Data Analysis XII*, pages 273–284, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

References III

P. Schäfer and M. Högqvist.
Sfa: A symbolic fourier approximation and index for similarity search in high dimensional datasets.
In *Proceedings of the 15th International Conference on Extending Database Technology*, EDBT '12, page 516–527. Association for Computing Machinery, 2012.

C. Truong, R. Barrois-Müller, T. Moreau, C. Provost, A. Vienne-Jumeau, A. Moreau, P.-P. Vidal, N. Vayatis, S. Buffat, A. Yelnik, D. Ricard, and L. Oudre.
A Data Set for the Study of Human Locomotion with Inertial Measurements Units.
Image Processing On Line, 9:381–390, 2019.
https://doi.org/10.5201/ipol.2019.265.

C. Truong, L. Oudre, and N. Vayatis.
Selective review of offline change point detection methods.
Signal Processing, 167:107299, 2020.