Time series machine learning with shapelets

Tony Bagnall & ...

A Hands-on Introduction to Time Series Classification and Regression

26/08/2024

Table of Contents

Notation one-stop shop

Before diving in, let us define some notations :

- a **time series** of length m and of dimension d as a vector $X = [\mathbf{x}_1, \dots, \mathbf{x}_m]$ with $\mathbf{x}_j \in \mathbb{R}^d$.
- a **time series dataset** $\mathcal{X} = [X_1, \dots, X_n]$, this is equivalent to an array 3D X of shape : (n_samples, n_channels, n_timepoints)
- an **annotation vector** $Y = [y_1, \dots, y_n]$. For example, in a regression context, $y_i \in \mathbb{R}$ would represent the label of X_i.

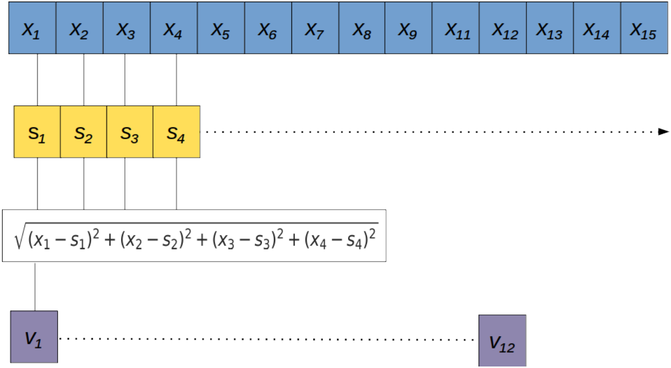
For simplicity, we'll consider time series with a fixed frequency. Additionally, we use:

- a blue color when talking about time series
- a yellow color when talking about shapelets
- a pink color when talking about annotation vectors
- a purple color when talking about distance vectors

What are shapelets ?

A shapelet [4] is a subsequence $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_l\} \in \mathbb{R}^{l,d}$ representing a pattern of interest. The similarity between S and X is quantified using a distance vector V. Example with $d = 1$ (univariate):

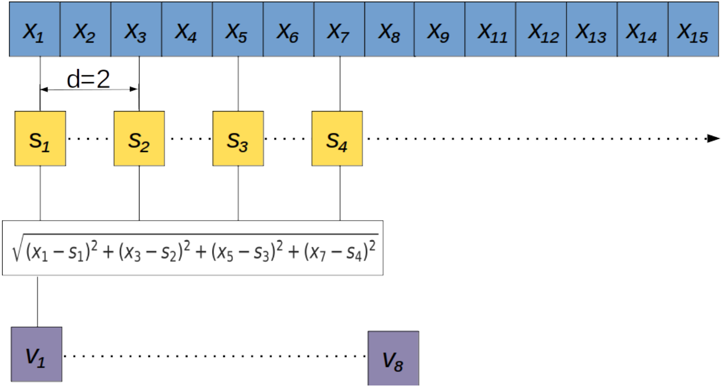
- $dist(X, S) = [v_1, \ldots, v_{m-(l-1)}]$
- with $v_i = \sqrt{\sum_{j=1}^{l}(s_j - x_{i+(j-1)})^2}$. (or any distance function)



What are (dilated) shapelets ?

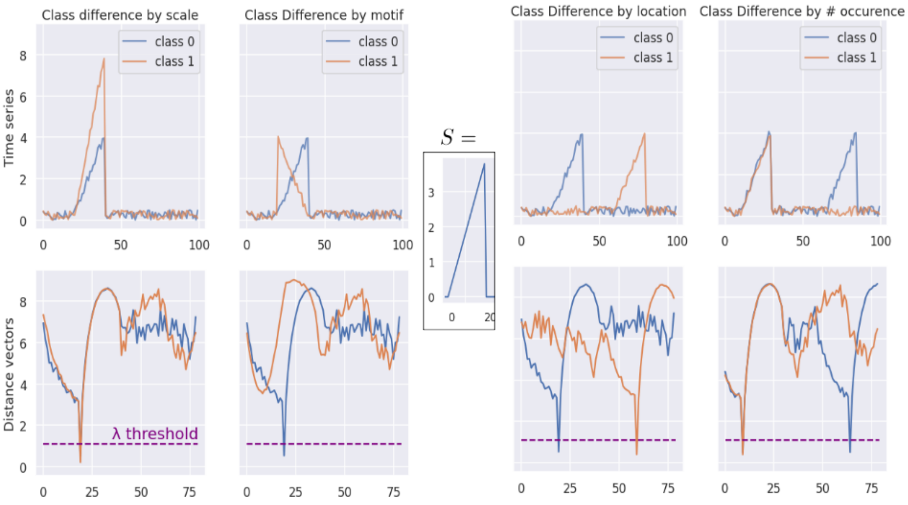
In [1], a parameter d is added to S. The intention is to :

- Allow matching non-contiguous, cyclic patterns (e.g. heartbeats)
- Increase receptive field.



What do we do with the distance vector ?

We want to identify discriminative characteristics between S and X:



What do we do with the distance vector ?

The following features are extracted to target these discriminative characteristics :

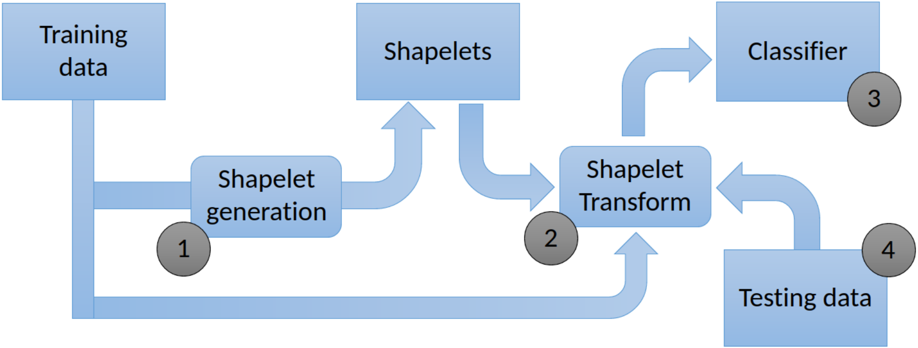
- **min** $d(X, S)$: the feature most shapelet-based methods use. It allows discriminating by **presence or absence of a pattern**.
- **argmin** $d(X, S)$: first used in [2] for shapelets, this allows to discriminate by **pattern location**.
- **Shapelet occurrence** [1]: $\sum_i I(v_i < \lambda)$, with I the identity function and a threshold λ, which is a shapelet parameter. It allows discriminating by **number of λ-close occurrence of the pattern**.

Standardizing S and every subsequence $[x_i, \ldots, x_{i+(l-1)}]$ independently makes v_i, and thus $d(X, S)$ (and the extracted features), **scale invariant**. Not standardizing makes them **scale sensitive**.

Shapelet-based classification and regression

All Shapelet classifiers or regressors use the same methodology, given $\mathcal{X} = \{X_1, \ldots, X_n\}$ and $Y = \{y_1, \ldots, y_n\}$:

1. Generate $\mathcal{S} = \{S_1, \ldots, S_k\}$ the set of Shapelets
2. Extract a matrix of features M, from $\mathcal{X}$ using $\mathcal{S}$
3. Learn a classifier based on M and Y



How do we generate Shapelets ?

We can distinguish three approaches :

- **Exhaustive search** : extract all possible shapelet candidates and evaluate their quality to extract the best ones (information gain, F-statistic, . . .)
- **Heuristic-based search**: random or semi-random extraction from the training data, statistical or rule-based methods for reducing the number of shapelet candidates.
- **Shapelet generation**: Use of optimization methods (gradient descent, evolutionary algorithm) to generate shapelet values instead of extracting them from the input.

What is the Shapelet Transform ?

Consider $\mathcal{X} = \{X_1, \ldots, X_n\}$ and $\mathcal{S} = \{S_1, \ldots, S_k\}$. The Shapelet Transform [3] algorithm is the same for all shapelet methods:

1. Create a matrix M of size (n, k)
2. $\forall i \in [1, n]$, $\forall j \in [1, k]$, and extract features from $d(X_i, S_j)$

For example, if extracting the minimum value :

$$M = \begin{bmatrix} \min d(X_1, S_1) & \ldots & \min d(X_n, S_1) \\ \vdots & \ddots & \vdots \\ \min d(X_1, S_k) & \ldots & \min d(X_n, S_k) \end{bmatrix}$$

A column M_i of M describes X_i by the features extracted using $\mathcal{S}$. The argmin can also be extracted, giving a matrix M of size $(n, 2k)$.

References I

Antoine Guillaume, Christel Vrain, and Wael Elloumi.
Random dilated shapelet transform: A new approach for time series shapelets.
In *Proceedings of the 3rd International Conference on Pattern Recognition and Artificial Intelligence*. Springer, 2022.

Mael Guillemé, Simon Malinowski, Romain Tavenard, and Xavier Renard.
Localized random shapelets.
In Vincent Lemaire, Simon Malinowski, Anthony Bagnall, Alexis Bondu, Thomas Guyet, and Romain Tavenard, editors, *Advanced Analytics and Learning on Temporal Data*, pages 85–97, Cham, 2020. Springer International Publishing.

References II

Jason Lines, Luke M. Davis, Jon Hills, and Anthony Bagnall.
A shapelet transform for time series classification.
pages 289–297, 2012.

Lexiang Ye and Eamonn Keogh.
Time series shapelets: A new primitive for data mining.
In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 947–956, New York, NY, USA, 2009. Association for Computing Machinery.