

Using sparse and interpretable features for hydrograph classification

J. Pelamatti^{1,3}, N. Bousquet^{1,3}, I. Seydi¹, F. Barjot², E. Paquet²

¹EDF R&D

²EDF DTG

³SINCLAIR AI Lab

1er Congrès Eau & IA, Grenoble

06/03/2026

Context

Explainable machine learning

It is more and more common to require machine learning models to be as interpretable/explainable as possible

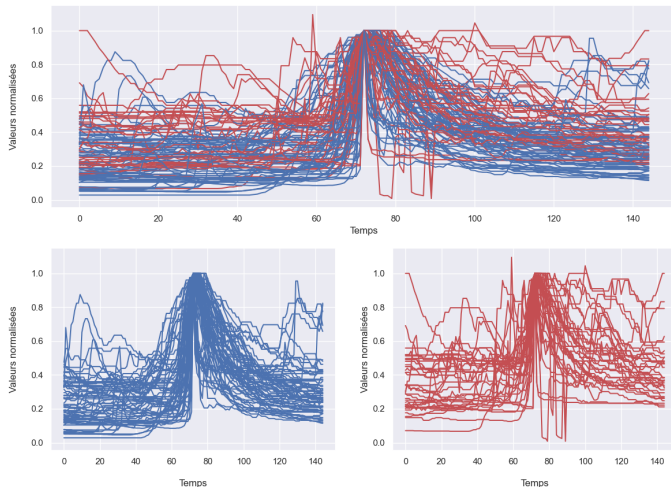
Various purposes:

- ▶ Critical applications
- ▶ Privacy safety
- ▶ Risk averse decision making, ...

Different approaches exist in order to ensure explainability and interpretability

- ▶ **We focus on creating inherently interpretable and independent features by combining shapelet transform and HSIC-Lasso feature selection**

Application : Flood hydrograph classification



- ▶ ~ 100000 flood hydrographs
- ▶ ~ 70 years span
- ▶ Measured for rivers of various sizes
- ▶ Presence of unrealistic measures (sensor errors, human errors, ...) $\rightarrow$ labeled data
- ▶ Can we automate the cleaning of new measurements with a classifier ?

SHAPELETS!

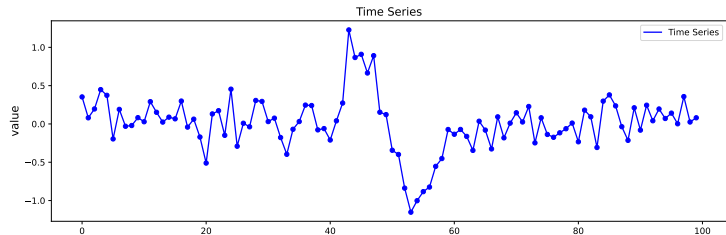
Shapelet transform of time series data

Underlying idea

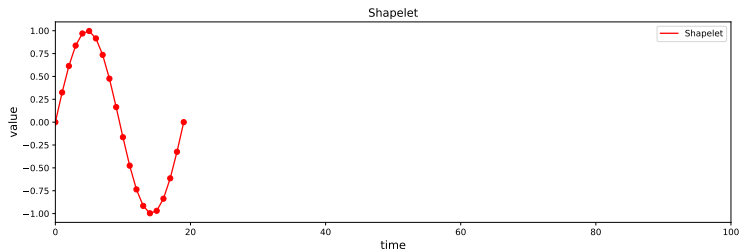
- ▶ Identifying discriminative time series patterns
- ▶ Characterizing each time series as a function of the minimum distance to these patterns
- ▶ Training a machine learning model on this transformed data set

Shapelet definition

- Time series
 $T = [t_1, \dots, t_m]$,
 $t_i \in \mathbb{R}$



- Shapelet
 $S = [s_1, \dots, s_l]$,
 $s_i \in \mathbb{R}$, $l < m$



Shapelet definition

The **projection** of a time series onto a shapelet can be computed as the minimum distance between S and any subsection of T of the same length:

$$\begin{aligned}
 f_S(T) &:= \mathbb{R}^m \rightarrow \mathbb{R} \\
 &= \min_{i=0, \dots, m-l} \text{dist}(T_{i,i+l}, S) \\
 &:= \min_{i=0, \dots, m-l} \sum_{j=1}^l \sqrt{(t_{i+j} - s_j)^2}
 \end{aligned}$$

In most cases, the two subseries are z-normalized before the computation

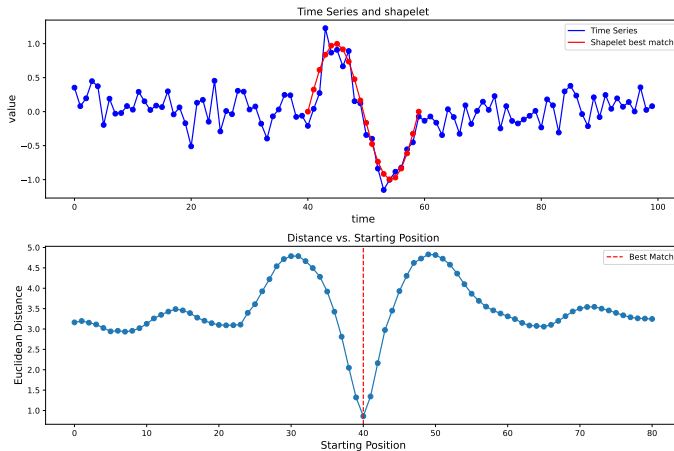
→ this distance becomes *amplitude invariant*

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



Shapelet definition

In other words, the minimum distance between the shapelet and a **sliding window** on the time-series



Shapelet definition

- ▶ Given a set of p shapelets $\mathbf{S} = [S_1, \dots, S_p]$, we can transform the m dimensional time series into a p -dimensional vector

$$\begin{aligned} f_{\mathbf{S}}(T) &:= \mathbb{R}^m \rightarrow \mathbb{R}^p \\ &:= [f_{S_1}(T), \dots, f_{S_p}(T)] \end{aligned}$$

- ▶ We then obtain a transformed data set

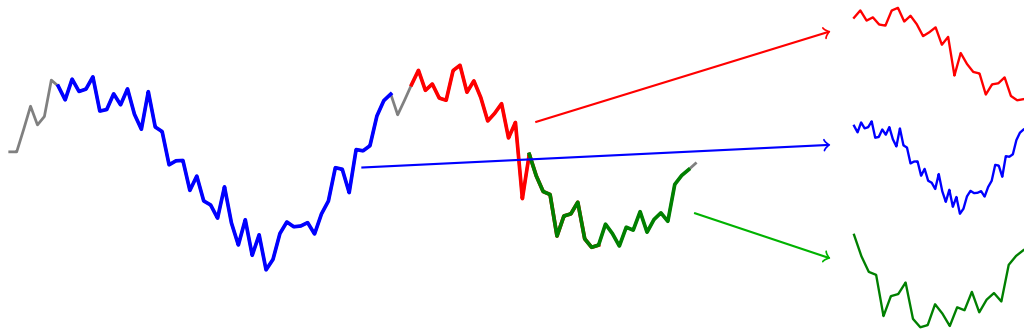
$$\mathbf{D}_{\mathbf{S}} = f_{\mathbf{S}}(\mathbf{T}), \in \mathbb{R}^{n \times p}$$

- ▶ Standard machine learning models can be trained on the p -dimensional data set $\{D_{i,\mathbf{S}}, y_i\}, i = 1, \dots, n$

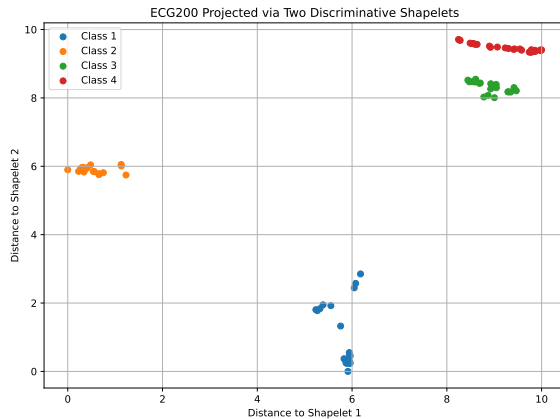
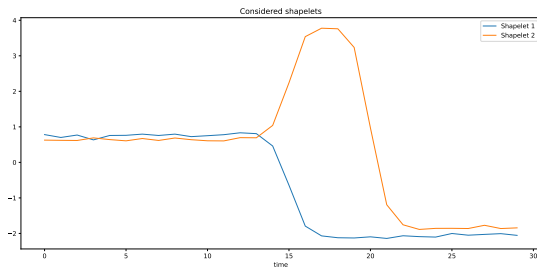
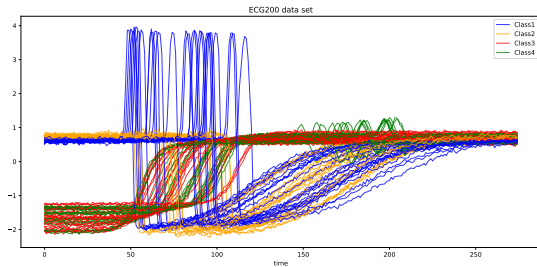
Shapelet definition

Selecting candidates shapelets

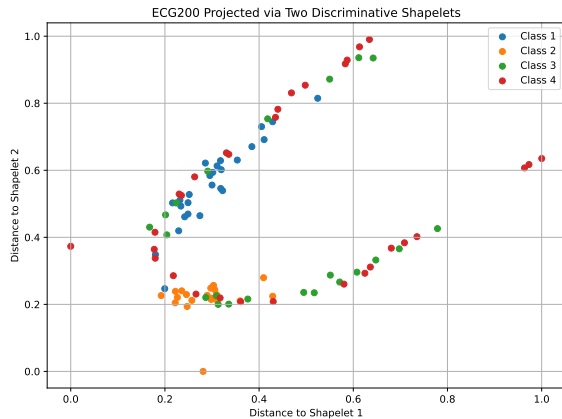
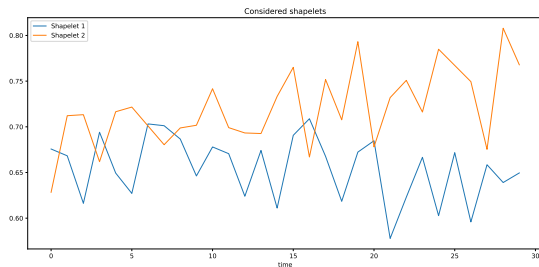
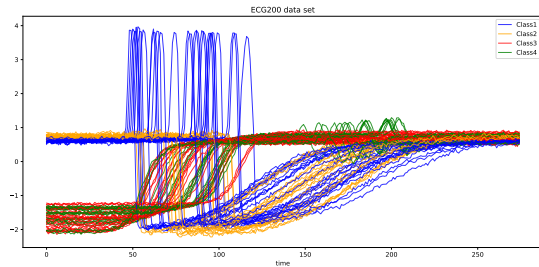
- ▶ A discriminative pattern is most likely to be present within the time series data set
- ▶ A common practice is to **randomly subsample** $N < N_{\max}$ candidates from the n time series



Shapelet definition



Shapelet definition



Shapelet definition

Selecting optimal shapelets

- ▶ We look for a limited set shapelets that generate the most *discriminative* feature set
- ▶ Ideally, optimal shapelets should also **carry physical meaning**
- ▶ *Standard* approach: the information gain

Some limitations

- ▶ Considerable computational cost & poor scaling
- ▶ Different shapelets may result in very similar transformations → the transformed data may present highly correlated dimensions

We look at the possibility of using HSIC-Lasso feature selection in order to identify informative and sparse shapelet transformations

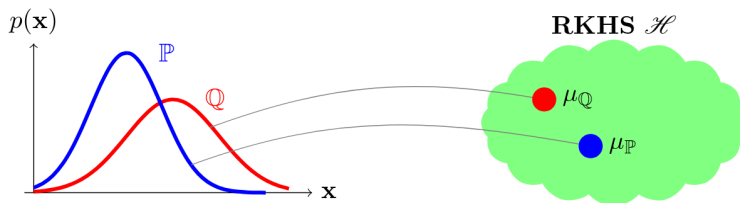
HSIC!

HSIC & stuff

- Quality of a shapelet $\leftrightarrow$ dependency between D_S and Y by comparing P_{Y,D_S} and $P_Y P_{D_S}$



$$\text{HSIC}(D_S, Y) := \|\mu[P_{Y,D_S}] - \mu[P_Y P_{D_S}]\|^2$$



[1] Image taken from Muandet, Krikamol, et al. "Kernel mean embedding of distributions: A review and beyond." (2017)

[2] A. Gretton et al. "Kernel Methods for Measuring Independence". In: Journal of Machine Learning Research 6 (2005), pp. 2075–2129.

HSIC & stuff

- ▶ $\text{HSIC}(D_S, Y) = 0 \Leftrightarrow D_S \perp Y$ Large $\text{HSIC}(D_S, Y)$ value $\Leftrightarrow$ strong dependence
- ▶ By the magic of RKHS, we have an estimator based on V-statistics

$$\widehat{\text{HSIC}}(D_S, Y) = \frac{1}{n^2} \text{Tr}(L_{D_S} H L_Y H)$$

- ▶ Only requires a data set and the computation of Gram and centering matrices
- ▶ Appropriate kernels can be used for continuous inputs and categorical outputs
 - ▶ Easily adapted to classification problems

HSIC-Lasso feature selection

We search for optimal shapelets as:

$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2} HSIC(y, y) - \sum_{i=1}^p \alpha_i HSIC(D_{S_i}, y) + \frac{1}{2} \sum_{i,j=1}^p \alpha_i \alpha_j HSIC(D_{S_i}, D_{S_j}) + \lambda \|\alpha\|_1$$

s.t., $\alpha_1, \dots, \alpha_p \geq 0$

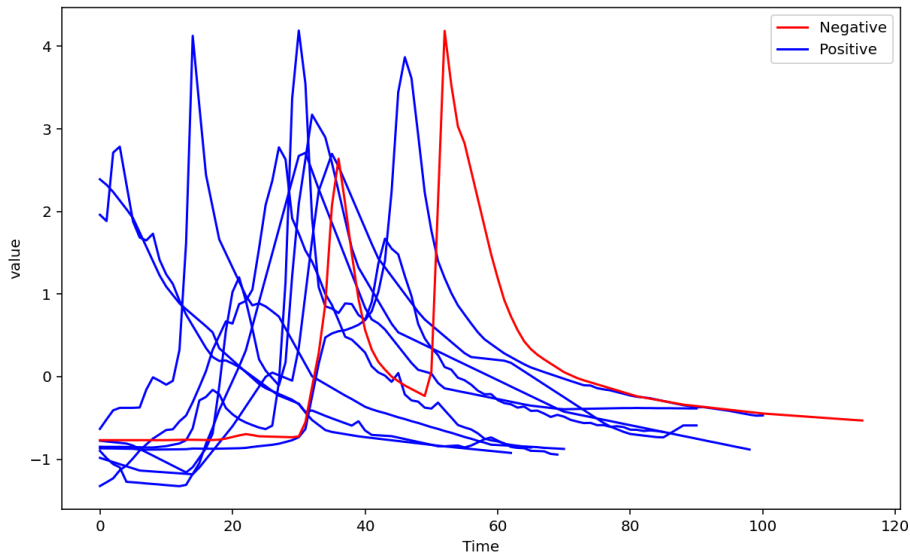
λ can be chosen so as to obtain a user-defined number of non-zero coefficients

- ▶ Discriminative features
- ▶ Independent features

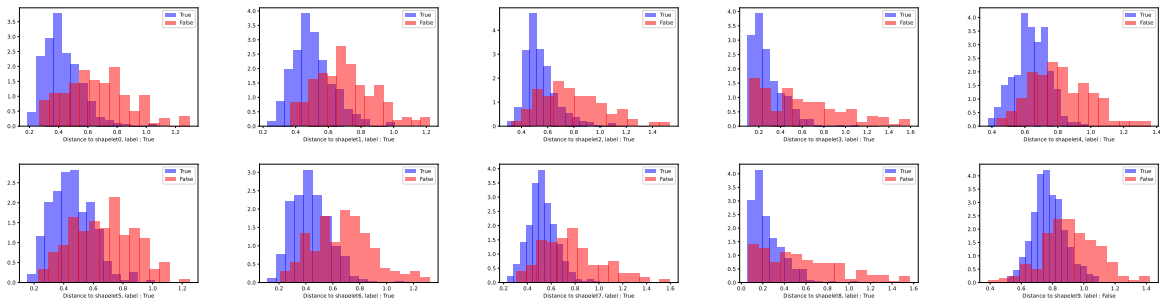
Flood analysis

- ▶ ~ 2500 time series
- ▶ 150 time steps
- ▶ 2 classes
- ▶ 10 selected shapelets among 5000 candidates
- ▶ Standard random forests are trained on the transformed data set
- ▶ We obtain a 82 – 84% classification accuracy
- ▶ Not too far from the labeling error rate
- ▶ BNo other available features have been used by the classifier, performances could be further improved

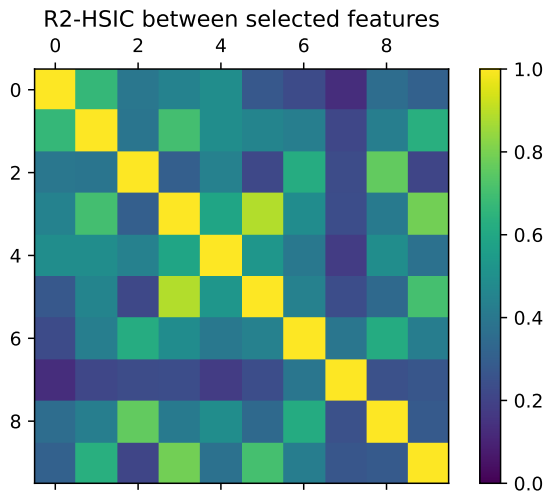
Numerical results: example of identified shapelets



Numerical results: density of the two classes on the various shapelet projections



Numerical results : dependence between selected features



Conclusions & Perspectives

- ▶ HSIC-Lasso can be used in the context of shapelet selection in order to perform high dimensional feature selection and identify **discriminative** and **independent** transformations of time series
- ▶ Better accuracy and computation time if compared to the original selection criteria (**not state of the art!**)
- ▶ The selected projections are sparse and (nearly) independent
→ more interpretable resulting ML models

Conclusions & Perspectives

BUT

- ▶ Need for a larger numerical benchmark, & comparison with other state of the art methods
- ▶ This approach should be coupled with existing speed-up techniques & alternative shapelet generation schemes (e.g., optimization)
- ▶ A binary kernel is used for the label output, more complex kernels should be considered in the case of multi-class classification
- ▶ HSIC-based clustering for automated labeling ?



"That's all Folks!"

Références I

- [Hil+14] Jon Hills et al. "Classification of time series by shapelet transformation". In: *Data mining and knowledge discovery* 28 (2014), pp. 851–881.
- [Gre+05] Arthur Gretton et al. "Kernel methods for measuring independence.". In: *Journal of Machine Learning Research* 6.12 (2005).
- [Mua+17] Krikamol Muandet et al. "Kernel mean embedding of distributions: A review and beyond". In: *Foundations and Trends® in Machine Learning* 10.1-2 (2017), pp. 1–141.
- [Yam+14] Makoto Yamada et al. "High-dimensional feature selection by feature-wise kernelized lasso". In: *Neural computation* 26.1 (2014), pp. 185–207.
- [YK09] 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*. 2009, pp. 947–956.