Time Series Shapelets

Jingchang Liu

August 25, 2019

University of Science and Technology of China

Table of Contents

Introduction to Time Series Shapelets

Introduction

- Shapelets are time series subsequences which are in some sense maximally representative of a class.
- Traditional represents may redundant.
- Traditional represents often have distortions.

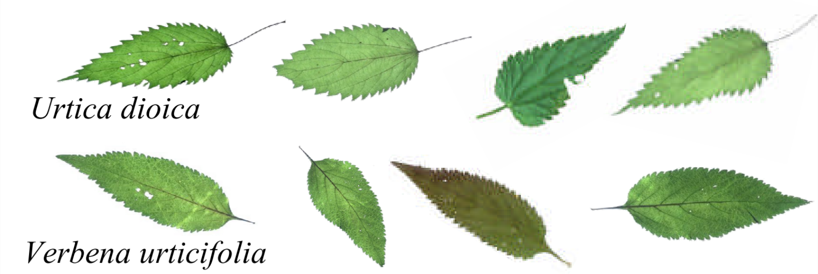

Figure 1: Samples of leaves from two species. Note that several leaves have the insect-bite damage

Shapelets

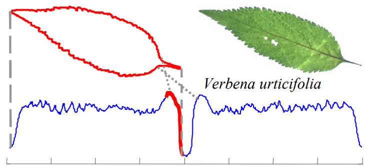

Figure 2: A shape can be converted into a one dimensional time series representation.

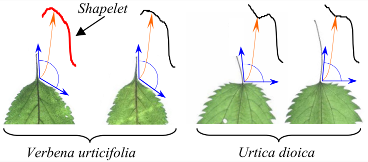

Figure 3: the shapelet has discovered that the defining difference between the two species is that Urtica dioica has a stem that connects to the leaf at almost 90 degrees, whereas the stem of Verbena urticifolia connects to the leaf at a much shallower angle.

Advantages

- Shapelets can provide interpretable results, which may help domain practitioners better understand their data
- Shapelets can be significantly more accurate/robust on some datasets. This is because they are local features.
- Shapelets can be significantly faster at classification than existing state-of-the-art approaches.

Brute-Force Algorithm

```
FindingShapeletBF (dataset D, MAXLEN, MINLEN)
candidates ← GenerateCandidates(D, MAXLEN, MINLEN)
bsf_gain ← 0
For each S in candidates
    gain ← CheckCandidate(D, S)
    If gain > bsf_gain
        bsf_gain ← gain
        bsf_shapelet ← S
    EndIf
EndFor
Return bsf_shapelet
```

Figure 4: Brute force algorithm for finding shapelet.

Learning Time-Series Shapelets

Framework

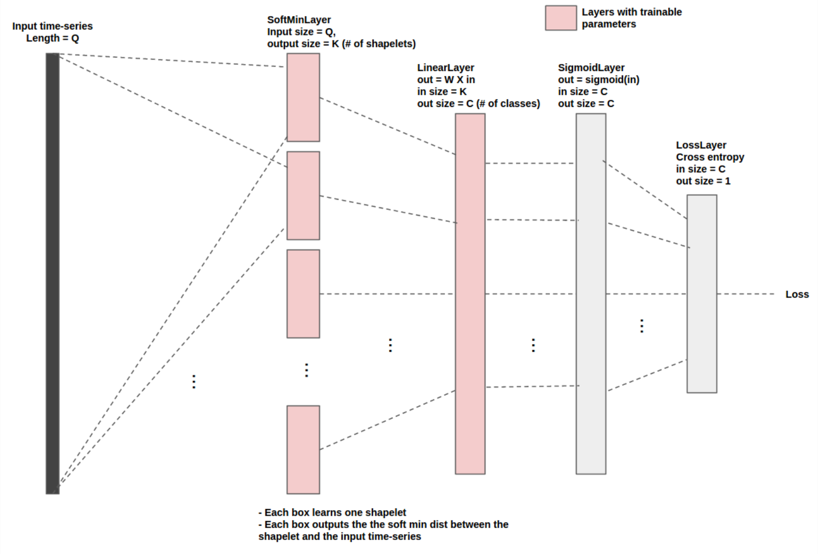

Figure 5: Framework of learning time-series shapelets.

Definitions and Notations

1. Time-Series Dataset. I training instances, each contains Q-many ordered values. Dataset is defined as $T^{I \times Q}$, whilethe series target is noticed as $Y \in \{0, 1\}$.
2. Sliding window segment of length L is an ordered sub-seq of a series $(T_{i,j}, \ldots, T_{i,j+L-1})$.
3. A shapelet of length L is simply an ordered sequence of values from a data structure perspective. K-most informative shapelets are denoted as $S \in \mathcal{R}^{K \times L}$.
4. The distance between k-th shapelets S_k and i-th series T_i is defined as the min distance $M_{i,k}$ among the distances between s_k and each segment j of T_i:

$$M_{i,k} = \min_{j=1,\ldots,J} \frac{1}{L} \sum_{l=1}^{L} (T_{i,j+l-1} - S_{k,l})^2 . \tag{1}$$

Learning Model

Prediction of *i*-th series
Since the minimum distance $M_k,\ k = 1, \ldots, K$ are new predictors, target of i-th series can be predicted as

$$\hat{Y}_i = W_0 + \sum_{k=1}^{K} M_{i,k} W_k, \quad \forall\, i \in \{1, \ldots, I\}. \tag{2}$$

Sigmoid function

$$\sigma(\hat{Y}_i) = \frac{1}{1 + e^{-\hat{Y}_i}}, \quad i \in \{1, \ldots, I\}. \tag{3}$$

Learning Model

Loss function between true target Y_i and predicted ones $\hat{Y}_i$

$$L(Y_i, \hat{Y}_i) = -Y_i \log(\sigma(\hat{Y}_i)) - (1 - Y_i) \log(1 - \sigma(\hat{Y}_i)), \quad i \in \{1, \ldots, I\}. \tag{4}$$

Regularized Objective Function

$$\arg\min_{S,W} \mathcal{F}(S, W) = \arg\min_{S,W} \sum_{I=1}^{I} L(Y_i, \hat{Y}_i) + \lambda_W \|W\|^2, \tag{5}$$

where S means shapelets, and W means their weight.

Shapelets are learnt

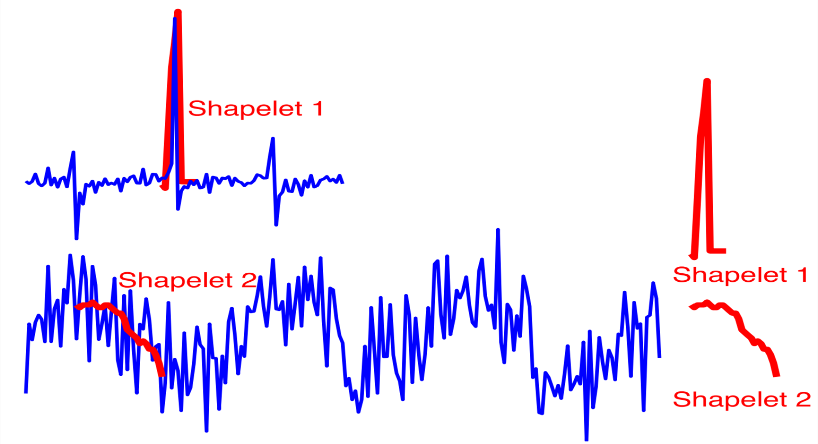

Figure 6: We can observe that the learnt shapelets may differ from all candidate segments and thus are robust to noise

Framework

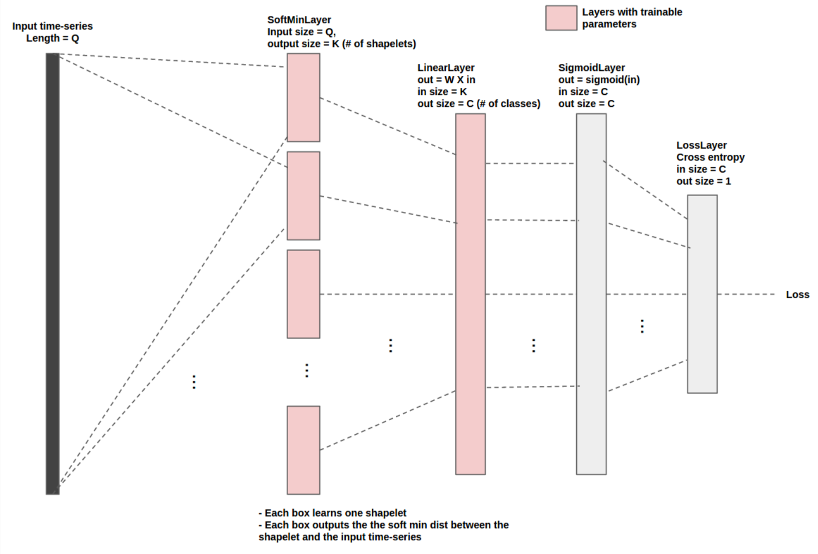

Figure 7: Framework of learning time-series shapelets.

Learning algorithm

Require: $T \in \mathbb{R}^{I \times Q}$, Number of Shapelets K, Length of a shapelet L, Regularization λ_W, Learning Rate η, Number of iterations: maxIter
Ensure: Shapelets $S \in \mathbb{R}^{K \times L}$, Classification weights $W \in \mathbb{R}^K$, Bias $W_0 \in \mathbb{R}$

```
for iteration=1, ..., maxIter do
    for i = 1, ..., I do
        W_0 ← W_0 − η ∂F_i/∂W_0
        for k = 1, ..., K do
            W_k ← W_k − η ∂F_i/∂W_k
            for L = 1, ..., L do
                S_{k,l} ← S_{k,l} − η ∂F_i/∂S_{k,l}
return S, W, W_0
```

Figure 8: Learning time-series shapelets.

But $M_{i,k} = \min_{j=1,}\{\cdots\}$'s gradient is hard to be handled!

Differentiable Minimum Function

- Distance between the j-th segment of series i and the k-th shapelet

$$D_{i,k,j} := \frac{1}{L} \sum_{l=1}^{L} \left(T_{i,j+l-1} - S_{k,l} \right)^2 , \tag{6}$$

- Soft minimum function

$$M_{i,k} \approx \hat{M}_{i,k} = \frac{\sum_{j=1}^{J} D_{i,k,j} e^{\alpha D_{i,k,j}}}{\sum_{j=1}^{J} e^{\alpha D_{i,k,j}}}, \tag{7}$$

  The smooth approximation of the minimum function, allows only the minimum segment to contribute for $\alpha \to -\infty$.

Gradients for Shapelets

$$\frac{\partial \mathcal{F}_i}{\partial S_{k,l}} = \frac{\partial L(Y_i, \hat{Y}_i)}{\partial \hat{Y}_i} \cdot \frac{\partial \hat{Y}_i}{\partial \hat{M}_{i,k}} \cdot \sum_{j=1}^{J} \frac{\partial \hat{M}_{i,k}}{\partial D_{i,k,j}} \cdot \frac{\partial D_{i,k,j}}{\partial S_{k,l}}, \tag{8}$$

where

$$\frac{\partial L(Y_i, \hat{Y}_i)}{\partial \hat{Y}_i} = -(Y_i - \sigma(\hat{Y}_i)), \tag{9}$$

$$\frac{\partial \hat{Y}_i}{\partial \hat{M}_{i,k}} = W_k, \tag{10}$$

$$\frac{\partial \hat{M}_{i,k}}{\partial D_{i,k,j}} = \frac{e^{\alpha D_{i,k,j}(1+\alpha(D_{i,k,j}-\hat{M}_{i,k}))}}{\sum_{j'=1}^{J} e^{\alpha D_{i,k,j'}}}, \tag{11}$$

and

$$\frac{\partial D_{i,k,j}}{\partial S_{k,l}} = \frac{2}{L}(S_{k,l} - T_{i,j+l-1}). \tag{12}$$

Gradients for Weights

$$\frac{\partial \mathcal{F}_i}{\partial W_k} = -(Y_i - \sigma(\hat{Y}_i))\hat{M}_{i,k} + \frac{2\lambda_W}{I} W_k \tag{13}$$

and

$$\frac{\partial \mathcal{F}_i}{\partial W_0} = -(Y_i - \sigma(\hat{Y}_i)). \tag{14}$$

Fused Lasso Generalized eigenvector method

Optimization Problem

$$\min_{v} v^{\mathrm{T}} C_2 v + \alpha_1 \|Dv\|_2 + \alpha_2 \|v\|_1, \ s.t. \ v^{\mathrm{T}} C_1 v = 1, \tag{15}$$

where D is a matrix such that $D_{i,i} = 1$, $D_{i,i+1} = -1$ and $D_{i,j} = 0$ otherwise, so $\|Dv\|_1 = \sum_i |v_i - v_{i-1}|$.

Opt
Rewrite (**??**) as

$$\begin{aligned} \min_{v,y,z} \quad &= \quad v^{\mathrm{T}} C_2 v + \alpha_1 \|z\|_1 + \alpha_2 \|y\|_1 \\ \text{s.t.} \quad & \quad Dv = z, \ v = y, \ v^{\mathrm{T}} C_1 v = 1, \end{aligned}$$

then use ADMM to solve it.

obtained v

$$v = [0, \ldots, 0, v_{s_1}, \ldots, v_{e_1}, 0, \ldots, 0, v_{s_B}, \ldots, v_{e_B}, 0, \ldots, 0]. \tag{16}$$

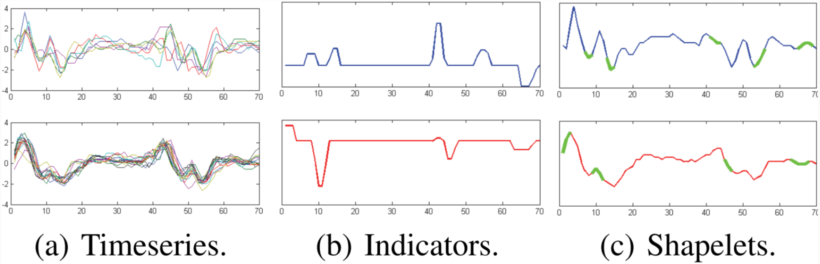

Figure 9: (a) Example timeseries from the two classes of SonyAIBORobotSurface; (b) Learnt shapelet indicator vectors; (c) Example shapelets.

Experiments

data set	IG	KW	FS	FSH	SD	IGSVM	LTS	UFS	FLAG
Adiac	29.9 (6)	26.6 (7)	15.6 (9)	57.5 (3)	52.2 (4)	23.5 (8)	51.9 (5)	69.8 (2)	**75.2 (1)**
Beef	50.0 (6)	33.3 (9)	56.7 (5)	50.0 (6)	50.0 (6)	**90.0 (1)**	76.7 (3)	66.7 (4)	83.3 (2)
Chlorine.	58.8 (5)	52.0 (9)	53.5 (8)	58.8 (5)	59.6 (4)	57.1 (7)	73.0 (3)	73.8 (2)	**76.0 (1)**
Coffee	96.4 (5)	85.7 (9)	**100.0 (1)**	92.9 (8)	96.4 (5)	**100.0 (1)**	**100.0 (1)**	96.4 (5)	**100.0 (1)**
Diatom.	76.5 (7)	62.1 (9)	76.5 (7)	87.3 (5)	86.6 (6)	93.1 (4)	94.2 (3)	95.8 (2)	**96.4 (1)**
DP_Little	- (8)	- (8)	- (8)	60.6 (5)	55.7 (6)	66.6 (4)	**73.4 (1)**	67.4 (3)	68.3 (2)
DP_Middle	- (8)	- (8)	- (8)	58.8 (5)	55.3 (6)	69.5 (3)	**74.1 (1)**	66.5 (4)	71.3 (2)
DP_Thumb	- (8)	- (8)	- (8)	63.4 (5)	54.4 (6)	69.6 (3)	**75.2 (1)**	68.5 (4)	70.5 (2)
ECGFiveDays	77.5 (9)	87.2 (8)	99.0 (4)	99.8 (3)	91.5 (7)	99.0 (4)	**100.0 (1)**	**100.0 (1)**	92.0 (6)
FaceFour	84.0 (6)	44.3 (8)	75.0 (9)	92.0 (4)	83.0 (7)	**97.7 (1)**	94.3 (2)	93.2 (3)	90.9 (5)
Gun_Point	89.3 (9)	94.0 (6)	95.3 (5)	94.0 (6)	93.1 (8)	**100.0 (1)**	99.6 (2)	98.7 (3)	96.7 (4)
ItalyPower.	89.2 (8)	91.0 (6)	93.1 (5)	91.0 (6)	88.0 (9)	93.7 (4)	**95.8 (1)**	94.0 (3)	94.6 (2)
Lightning7	49.3 (7)	48.0 (8)	41.1 (9)	65.2 (4)	65.2 (4)	63.0 (6)	**79.0 (1)**	68.5 (3)	76.7 (2)
MedicalImages	48.8 (8)	47.1 (9)	50.8 (7)	64.7 (5)	66.0 (4)	52.2 (6)	71.3 (2)	71.1 (3)	**71.4 (1)**
MoteStrain	82.5 (8)	84.0 (5)	84.0 (5)	83.8 (7)	78.3 (9)	88.7 (3)	**90.0 (1)**	87.2 (4)	88.8 (2)
MP_Little	- (8)	- (8)	- (8)	56.9 (6)	62.7 (5)	70.7 (3)	**74.3 (1)**	71.7 (2)	69.3 (4)
MP_Middle	- (8)	- (8)	- (8)	60.3 (6)	64.5 (5)	76.9 (2)	**77.5 (1)**	74.8 (4)	75.0 (3)
Otoliths	67.2 (1)	60.9 (6)	57.8 (8)	60.9 (6)	64.1 (2)	64.1 (2)	59.4 (5)	57.8 (8)	64.1 (2)
PP_Little	- (8)	- (8)	- (8)	57.6 (5)	55.8 (6)	**72.1 (1)**	71.0 (2)	66.8 (4)	67.1 (3)
PP_Middle	- (8)	- (8)	- (8)	61.6 (5)	60.5 (6)	**75.9 (1)**	74.9 (3)	75.4 (2)	73.8 (4)
PP_Thumb	- (8)	- (8)	- (8)	55.8 (6)	61.8 (5)	**75.5 (1)**	70.5 (2)	67.2 (4)	67.4 (3)
Sony.	85.7 (5)	72.7 (8)	**95.3 (1)**	68.6 (9)	85.0 (6)	92.7 (3)	91.0 (4)	79.0 (7)	92.9 (2)
Symbols	78.4 (8)	55.7 (9)	80.1 (7)	92.4 (2)	86.5 (5)	84.6 (6)	**94.5 (1)**	88.8 (3)	87.5 (4)
SyntheticC.	94.3 (7)	90.0 (8)	95.7 (5)	94.7 (6)	98.3 (3)	87.3 (9)	97.3 (4)	**99.7 (1)**	**99.7 (1)**
Trace	98.0 (5)	94.0 (9)	**100.0 (1)**	**100.0 (1)**	96.0 (7)	98.0 (6)	**100.0 (1)**	96.0 (7)	99.0 (4)
TwoLeadECG	85.1 (7)	76.4 (9)	97.0 (4)	92.5 (5)	86.7 (6)	**100.0 (1)**	**100.0 (1)**	83.6 (8)	99.0 (3)
average rank	6.9	7.7	6.3	5.2	5.9	3.5	2.0	3.7	2.6

Figure 10: Testing accuracies (%) on the data sets.

Conclusions

Conclusions

- Time series shapelets is the most representative seq of series.
- Learning shapelets is the right direction,.
- There can be more improvement of learning shapelets.

Q & A