Time Series Data Mining Challenges

Jose A. Lozano

Basque Center for Applied Mathematics (BCAM)
University of the Basque Country UPV/EHU

EVIA, Sevilla 14 Junio, 2023
Outline of the presentation

1. Time Series Data Mining Activities
2. Clustering
3. (Early) Supervised Classification
4. Outlier/Anomaly Detection
5. Conclusions and Future Work
Outline of the presentation

1. Time Series Data Mining Activities
2. Clustering
3. (Early) Supervised Classification
4. Outlier/Anomaly Detection
5. Conclusions and Future Work
Time series are all around
Definition and main characteristics

Definition

A **time series** is an ordered sequence of pairs of finite length $L$:

$$TS = \{(t_i, x_i) | \ i = 1, ..., L\}, \quad (1)$$

where the timestamps $\{t_i\}_{i=1}^{L}$ take positive and ascending real values and the values of the time series $(x_i)$ take univariate or multivariate real values.

Main characteristics

- Temporal correlation
- High dimensionality
- Noise
Time Series Data Mining Challenges

Time Series Data Mining Activities

Time series forecasting

Examples

- Stock market prediction
- Temperature prediction
Time series data base: our object of study

- A set of time series (usually big)
- Different lengths
- Multidimensional
Time series clustering. Examples
Supervised classification of time series

TRAINING SET

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph1" /></td>
<td><img src="image2.png" alt="Graph2" /></td>
<td><img src="image3.png" alt="Graph3" /></td>
</tr>
</tbody>
</table>

ALGORITHM

CLASSIFIER

C2
Anomaly/outlier detection
Segmentation

A)

B)
Outline of the presentation

1. Time Series Data Mining Activities
2. Clustering
3. (Early) Supervised Classification
4. Outlier/Anomaly Detection
5. Conclusions and Future Work
Time series clustering. Examples
Differences with the classic clustering problem
Time series clustering: hierarchical, partitional

we need a DISTANCE
Distance between time series

Rigid Distance

Flexible Distance
Euclidean Distance (ED)

\[ D(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]

- Easy to compute ✅
- Only for series with the same distance ✅
- Does not consider the time ✅
- Sensitivity to noise ✅
- Requires the series to be sampled at the time stamps ✅
Dynamic Time Warping (DTW)

- Takes into account the ordered sequence (time)
- It can deal with series of different sizes
- Does not require to be sampled in the same time stamps
- Computationally expensive \( O(\min\{m, n\}^2) \)
Euclidean Distance vs Dynamic Time Warping

**EUCLIDEAN**

**DTW**
More on elastic distances

Cheap versions of dynamic time warping
More on elastic distances

Edit distances for real sequences (LCSS, EDR, ERP)

Mori et al., 2016, R journal
More on elastic distances

On-line versions (Oregi et al 2019, PR)

(a) Incremental DTW computation.

(b) Weights over the DTW lattice.
Represent each series by means of a set of features and calculate the distance between the features

Learn a **parametric model** for each series and calculate the distance between the parameters
Last remarks on distances between series

Remarks

- There is **no best distance** (no free lunch)
- Each problem requires a different distance
- The distance to be used needs to be in agreement with our knowledge about what is far and what is close
- Hint: try several distances

**Challenge:**
Design a method to the (semi)automatic selection of a distance (e.g. Mori et al. 2016, TKDE)
...back to clustering: problems with $k$-means

$k$-means

$k$-medoids
Remarks on clustering

- Recent papers on the computation of a *mean* series
- Alternate clustering methods: graph-based, spectral, model-based, ...
- Multivariate time series clustering
Outline of the presentation

1. Time Series Data Mining Activities
2. Clustering
3. (Early) Supervised Classification
4. Outlier/Anomaly Detection
5. Conclusions and Future Work
Supervised classification of time series

General-purpose classifiers

Specific TS classifiers
Each series is considered an instance
Each time stamp is considered a feature

<table>
<thead>
<tr>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$\ldots$</th>
<th>$t_n$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{11}$</td>
<td>$x_{12}$</td>
<td>$x_{13}$</td>
<td>$\ldots$</td>
<td>$x_{1n}$</td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_{21}$</td>
<td>$x_{22}$</td>
<td>$x_{23}$</td>
<td>$\ldots$</td>
<td>$x_{2n}$</td>
<td>$c_2$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$x_{m1}$</td>
<td>$x_{m2}$</td>
<td>$x_{m3}$</td>
<td>$\ldots$</td>
<td>$x_{mn}$</td>
<td>$c_2$</td>
</tr>
</tbody>
</table>
General-purpose classifiers

- Each series is considered an instance
- Each time stamp is considered a feature

<table>
<thead>
<tr>
<th></th>
<th>$t_2$</th>
<th>$t_1$</th>
<th>$t_3$</th>
<th>...</th>
<th>$t_n$</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{12}$</td>
<td>$x_{11}$</td>
<td>$x_{13}$</td>
<td>...</td>
<td>$x_{1n}$</td>
<td>$c_1$</td>
<td></td>
</tr>
<tr>
<td>$x_{22}$</td>
<td>$x_{21}$</td>
<td>$x_{23}$</td>
<td>...</td>
<td>$x_{2n}$</td>
<td>$c_2$</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$x_{m2}$</td>
<td>$x_{m1}$</td>
<td>$x_{m3}$</td>
<td>...</td>
<td>$x_{mn}$</td>
<td>$c_2$</td>
<td></td>
</tr>
</tbody>
</table>
General-purpose classifiers

- Each series is considered an instance
- Each time stamp is considered a feature

<table>
<thead>
<tr>
<th></th>
<th>$t_2$</th>
<th>$t_1$</th>
<th>$t_3$</th>
<th>$t_n$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{12}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$c_1$</td>
</tr>
<tr>
<td>$x_{22}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$c_2$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$x_{m2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$c_2$</td>
</tr>
</tbody>
</table>

**CHALLENGE**

When to use general-purpose and when time-series specific?
What is relevant in TSC?

PROBLEM I

PROBLEM II
What is relevant in TSC?

**PROBLEM I**

SHAPE

**PROBLEM II**

SHAPE
What is relevant in TSC?

PROBLEM I

SHAPE

PROBLEM II

LOCATION
A taxonomy of time series classification methods

Taxonomy

- Distance-based classifiers
- Model-based classifiers
- Feature-based classifiers
Taxonomy of distance-based TSC (Abanda et al. 2019, DAMI)

- **k-NN**
  - Distances are used combined with k-NN classifiers

- **Distance features**
  - Distances are used to obtain a new feature representation of the time series
  - **Global distance features**
    - Distances to other (global) series are used as features
  - **Local distance features**
    - Distances to local patterns of the series are used as features
  - **Embedded features**
    - Distances are used to embed the series into a vector space and obtain new features

- **Distance kernels**
  - Distances are used to obtain a kernel
  - **Indefinite distance kernels**
    - Distances are used to obtain indefinite kernels for time series
  - **Definite distance kernels**
    - Distances are used to obtain PSD kernels for time series
1-Nearest Neighbour (1-NN)

```
<table>
<thead>
<tr>
<th></th>
<th>C_1</th>
<th>C_2</th>
<th>C_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

\[(d_1 \quad d_2 \quad d_3 \quad d_4 \quad d_5 \quad d_6)\]  

MINIMUM DISTANCE  

\[C_2\]
1-Nearest Neighbour (1-NN)

- Easy to understand
- Better results with higher number of series
- Computational cost
- Challenge: What distance???

![Diagram showing 1-NN with distance calculation](image)

MINIMUM DISTANCE
Distance-based. Distance features. Global
Distance-based. Distance features. Global
Distance-based. Distance features. Global

- Any general-purpose algorithm could be applied
- It depends on the number of series in training set
- Computationally expensive
Time Series Data Mining Challenges
(Early) Supervised Classification

Distance-based. Distance features. Local

Shapelet 1  Shapelet 2  Shapelet 3

$L_{ij}$ could be distance or presence
- Computationally expensive ✓
- When the shapelets are relevant extremely good results ✓
- Easy to interpret ✓
Distance-based Embedding

Training set

<table>
<thead>
<tr>
<th>TS_1</th>
<th>C_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS_2</td>
<td>C_2</td>
</tr>
<tr>
<td>TS_3</td>
<td>C_3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>TS_n</td>
<td>C_n</td>
</tr>
</tbody>
</table>

Distance matrix

<table>
<thead>
<tr>
<th>TS_1</th>
<th>...</th>
<th>TS_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTW(TS_1,TS_1)</td>
<td>...</td>
<td>DTW(TS_n,TS_n)</td>
</tr>
<tr>
<td>C_1</td>
<td>...</td>
<td>C_n</td>
</tr>
</tbody>
</table>

Approximated distance matrix

<table>
<thead>
<tr>
<th>V_1</th>
<th>...</th>
<th>V_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED(v_1,v_1)</td>
<td>...</td>
<td>ED(v_n,v_n)</td>
</tr>
<tr>
<td>C_1</td>
<td>...</td>
<td>C_n</td>
</tr>
</tbody>
</table>

Embedded vectors

<table>
<thead>
<tr>
<th>V_1</th>
<th>C_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>V_2</td>
<td>C_2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>V_n</td>
<td>C_n</td>
</tr>
</tbody>
</table>

Classifier

DTW

(Early) Supervised Classification
(Early) Supervised Classification

Distance-based Embedding

- Many classifiers defined in Euclidean spaces
- Computational complexity
- Prediction

Training set:

<table>
<thead>
<tr>
<th>TS_1</th>
<th>C_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS_2</td>
<td>C_2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>TS_n</td>
<td>C_n</td>
</tr>
</tbody>
</table>

Distance matrix:

<table>
<thead>
<tr>
<th>TS_1</th>
<th>...</th>
<th>TS_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTW(TS_1, TS_1)</td>
<td>...</td>
<td>DTW(TS_n, TS_n)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Approximated distance matrix:

<table>
<thead>
<tr>
<th>V_1</th>
<th>...</th>
<th>V_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED(V_1, V_1)</td>
<td>...</td>
<td>ED(V_n, V_1)</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Embedded vectors:

<table>
<thead>
<tr>
<th>V_1</th>
<th>C_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>V_2</td>
<td>C_2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>V_n</td>
<td>C_n</td>
</tr>
</tbody>
</table>

Embedding process:

DTW → Embedding → Classifier
Distance Kernels

Definite (PSD) Kernel
- All the SVM machinery works
- Difficult to define/check

Indefinite
- Theoretical properties are lost
- Easy to define
- Some methods cannot be applied
Distance Kernels. Indefinite

Gaussian Distance Substitution Kernels

\[ GDS_d(x, x') = \exp \left( - \frac{d(x, x')^2}{\sigma^2} \right) \] where \( d = DTW, .. \)
Feature-based time series classification

- SERIES
- FEATURES
- CLASSIFIER

\[
\begin{array}{cccc}
  f_1^1 & f_2^1 & \cdots & f_k^1 \\
  f_1^2 & f_2^2 & \cdots & f_k^2 \\
  f_1^3 & f_2^3 & \cdots & f_k^3 \\
  f_1^4 & f_2^4 & \cdots & f_k^4 \\
  f_1^5 & f_2^5 & \cdots & f_k^5 \\
  f_1^6 & f_2^6 & \cdots & f_k^6 \\
\end{array}
\]
Feature-based time series classification
Feature-based time series classification

- Statistics: mean, variance
- Autorregresive coefficients
- Fourier coefficients
- Shift, trend, ...
- tsfeatures (Yang et al. 2015)
Feature-based time series classification

- Representation independent on the number of series
- Interpretable representation
- Challenge: what features to use?
Model-based time series classification

What is the most probable model?
Model-based time series classification

- Good results with an appropriate model ✓
- Choice of model ✓
- Existence of model ✓
Early time series classification

Examples

- Early activity recognition
- Early disease recognition in electrocardiograms
- Early detection of sepsis in newborn
- Early detection of failures in machines (predictive maintenance)
Early time series classification

**Balance between accuracy and earlyness**

**ALGORITHM**

**TRAINING SET**

<table>
<thead>
<tr>
<th></th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="blue" alt="Graph 1" /></td>
<td><img src="red" alt="Graph 2" /></td>
<td><img src="green" alt="Graph 3" /></td>
</tr>
<tr>
<td></td>
<td><img src="blue" alt="Graph 1" /></td>
<td><img src="red" alt="Graph 2" /></td>
<td><img src="green" alt="Graph 3" /></td>
</tr>
</tbody>
</table>

**EARLY CLASSIFIER**

- ?
- ?

**CLASSIFIER EARLY**

- ?

- Wait for more data

- C₂
Early time series classification (Mori et al 2017, DAMI, TNNLS)
Early time series classification

Time Series Data Mining Challenges

(Early) Supervised Classification
Early time series classification

Output BLUE class
Multivariate time series classification

CHALLENGE
Outline of the presentation

1. Time Series Data Mining Activities
2. Clustering
3. (Early) Supervised Classification
4. Outlier/Anomaly Detection
5. Conclusions and Future Work
Outlier vs Anomaly

Outliers meaning

- Event of interest
  - aim: Detect the outlier itself

- Unwanted data
  - aim: Data cleaning

- Improve the data quality for further analysis
Type of outlier: point outlier
Type of outlier: subsequence outlier
Type of outlier: series outlier
Outlier detection method: basic

\[ |X_t - \hat{X}_t| > \tau \]
Outlier detection method: basic

\[ |x_t - \hat{x}_t| > \tau \]

Median
Outlier detection method: basic

\[ |x_t - \hat{x}_t| > \tau \]

MAD
Outlier detection method: basic

\[ |x_t - \hat{x}_t| > \tau \]
An overview of outlier/anomaly detection

- **Input data**
  - Univariate time series
  - Multivariate time series

- **Outlier type**
  - Point
  - Subsequence
  - Time series

- **Detection method**
  - Univariate
  - Multivariate

- **Processing mode**
  - Offline
  - Online
Outline of the presentation

1. Time Series Data Mining Activities
2. Clustering
3. (Early) Supervised Classification
4. Outlier/Anomaly Detection
5. Conclusions and Future Work
Not too explored lands

Challenges

- Time series subset selection
- Learning in weakly environments: semi-supervised, multi-label, crowd learning
- Theoretical bounds on learning: assumptions on the generating model
Collaboration

- Usue Mori (UPV/EHU), Amaia Abanda (BCAM)
- Ane Blazquez (Ikerlan), Angel Conde (Ikerlan)
- Aritz Perez (BCAM), Izaskun Oregui (Tecnalia), Javier del Ser (Tecnalia)
- Josu Ircio (Ikerlan), Aizea Lojo (Ikerlan)