Harnessing the Power of Time Series Data

Soukaina Filali Boubrahimi
Agenda:

1. State of the Art Classification Methods
2. New DNN Methods for TSC
3. 1DCNN for Shapelet Mining and TSC
4. Application of TSC : SEP Prediction
5. Conclusion & Future Plans
Methods For Time Series Data Classification

Dictionary Based

...reduce the dimensionality of series by transforming them into representative words, then comparing the distribution of words...

Shapelet Based

Shapelet: TS subsequences with high discriminatory power

Interval Based

Interval based classifiers derive features from intervals of each time series.

Differential Distance Based

...are based on the first order differences of the series:

\[ a'_i = a_i - a_{i+1} \quad i = 1 \ldots m - 1 \]

Time Domain Based

k-Nearest Neighbor search classifier coupled with elastic/lockstep measures for shape based classification
Methods For Time Series Data Classification

**Dictionary Based**

...reduce the dimensionality of series by transforming them into representative words, then comparing the distribution of words...

**Shapelet Based**

Shapelet: TS subsequences with high discriminatory power

**Interval Based**

Interval based classifiers derive features from intervals of each time series.

**Differential Distance Based**

...are based on the first order differences of the series:

\[ a'_i = a_i - a_{i+1} \quad i = 1 \ldots m - 1 \]

**Time Domain Based**

k-Nearest Neighbor search classifier coupled with elastic/lockstep measures for shape based classification

**Deep Learning Methods**

Adjust state-of-the art DNN architectures for time series data input
MLP is the simplest and most traditional model. Consists of three-layered model of 500 neurons, that are fully connected to next layer neurons. Uses dropout method between layers. A layer block can be formalized as follows:

\[
\tilde{x} = f_{\text{dropout}, p}(x) \\
y = W \cdot \tilde{x} + b \\
h = \text{ReLU}(y)
\]

In practice, MLP is too simple for the TSC task.

Deep Learning Methods: FCN (Wang et al., 2016)

- CNN has revolutionized the field of computer vision for image classification.
- FCNN consists of three-layered fixed sized kernel model.
- A layer block can be formalized as follows:

\[
\begin{align*}
y &= W \otimes x + b \\
s &= BN(y) \\
h &= ReLU(s)
\end{align*}
\]

- FCN achieves best rank within DNN category.

Deep Learning Methods: ResNet (Wang et al., 2016)

- ResNet is the most complex model within DNN category.
- Consists of three residual blocks, in their turns consisting of 3 fixed-sized layers.
- A layer block can be formalized as follows:

\[
\begin{align*}
    h_1 &= Block_{k_1}(x) \\
    h_2 &= Block_{k_2}(h_1) \\
    h_3 &= Block_{k_3}(h_2) \\
    y &= h_3 + x \\
    \hat{h} &= ReLU(y)
\end{align*}
\]

- ResNet is too deep for most TSC datasets
Learned Shapelets (LS) (Grabocka, 2014)

- Evolution of 2 learned shapelets with progressing iterations with hyperparameters:
  \[ L = 40, \; \mu = 0.01, \; \lambda_w = 0.01, \; \alpha = -100, \; s1 = \text{---}, \; s2 = \text{---} \]

Limitation of the method => Considers only one shapelets length (determined during Cross-Validation)

In analogy with 2-D CNN for image model, we adjusted the architecture for time series data input.

- **Our hypothesis is**: “a dataset can have discriminative shapelets of different lengths all of which can contribute with independent information to the classification”

- 1DCNN has a variable size of kernels in terms of both numerosity and lengths.

- Three flavors of 1DCNN were considers: Fixed kernels (wide and long) and variable-length kernels.
1D-Convolutional Neural Network for TSC: Architecture

- 1DCNN consists of three layers:
  1. **Conv. layer**: Three parallel kernel blocks of: \( \{2.5\%, 12.5\%, 20\%\} \times \) original Time Series length.
  2. **Max-pooling layer**: Applies max pooling operation and concatenate all input
  3. **Fully Connected layer**:

- A layer block can be formalized as follows:

\[
\begin{align*}
    y &= W \odot x + b \\
    h &= \text{ReLU}(y) = \max(0, y) \\
    s &= \max(h)
\end{align*}
\]
Applying a one-dimensional convolution operation followed by a max pooling on the feature map is equivalent to finding the best shapelet alignment in the time series and dropping all other alignments.
Datasets

- 1DCNN was tested on the University of California Riverside which consists of 85 datasets from different domains and different characteristics \{time series length, number of examples, number of classes...\}

1D-Convolutional Neural Network for TSC : Mined Shapelets

- **1DCNN always** performs better than fixed kernel size architecture (deep and long kernels)

- 1DCNN achieves better accuracy levels than shapelets method and comparable results to FCN

- As a by-product of 1DCNN classifier, the 1D kernels are interpretable (represents mined shapelets) compared to FCN 2-D kernels.
1DCNN Summary

**Problem:** Typical DNN architectures are black boxes whose kernels lacks to model interpretable discriminative features in time series data.

**Approach:** Use a 1D CNN with 3 parallel layers whose kernels represent learned discriminative shapelets of different lengths (proportional to TS length).
Application of TSC: SEP Prediction

Problem
Prediction of SEP events >100 MeV given an X-Ray event

Goal
Extend the previous works by studying the proton-xray correlation and proton channels cross-correlations.

Solution
Use Vector AutoRegressive (VAR) model to produce features and a decision tree model for the classification.

## Example of SEP and NSEP Events

<table>
<thead>
<tr>
<th>Fast-Rising</th>
<th>Slow-Rising</th>
<th>Lack of SEP</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

SEP and NSEP Events Catalogs

Dataset

- We used a total of 94 events (47 SEP and 47 non-SEP)
- The time period of the study: [1997-2006]
- Missing Data: 7 SEP events had missing data in their P6 and P7 channels. We selected a corresponding 7 non-SEP events with missing data in the same channels.
- Sampling: 10-fold cross validation

VAR models (vector Autoregressive models) are used for multivariate time series.

The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.

The vector autoregressive model of order 1, VAR(1), is as follows:

$$P_{n,t,1} = \phi_{P_{n,x},1} X_{s,t-1,1} + \phi_{P_{n,xL},1} X_{L,t-1,1} + \sum_{i=6}^{11} \phi_{P_{n,P_{i},1}} P_{i,t-1,1}, \quad n \in [6,11]$$

Vector AutoRegression Model

- Explicitly the system of VAR(1) model is expressed as:

\[
P_{6t,1} = \phi_{P6_{-x},1} \times X_{St-1,1} + \ldots + \phi_{P6_{-P11,1}} \times P_{11t-1,1} + \alpha_{P6t,1} \tag{1}
\]

\[
P_{7t,1} = \phi_{P7_{-x},1} \times X_{St-1,1} + \ldots + \phi_{P7_{-P11,1}} \times P_{11t-1,1} + \alpha_{P7t,1} \tag{2}
\]

\[
P_{8t,1} = \phi_{P8_{-x},1} \times X_{St-1,1} + \ldots + \phi_{P8_{-P11,1}} \times P_{11t-1,1} + \alpha_{P8t,1} \tag{3}
\]

\[
P_{9t,1} = \phi_{P9_{-x},1} \times X_{St-1,1} + \ldots + \phi_{P9_{-P11,1}} \times P_{11t-1,1} + \alpha_{P9t,1} \tag{4}
\]

\[
P_{10t,1} = \phi_{P10_{-x},1} \times X_{St-1,1} + \ldots + \phi_{P10_{-P11,1}} \times P_{11t-1,1} + \alpha_{P10t,1} \tag{5}
\]

\[
P_{11t,1} = \phi_{P11_{-x},1} \times X_{St-1,1} + \ldots + \phi_{P11_{-P11,1}} \times P_{11t-1,1} + \alpha_{P11t,1} \tag{6}
\]

Vector AutoRegression Model

- One data point (SEP/NSEP event) is represented by the following vector:

\[
x = \begin{pmatrix}
\phi_{P6_{xs}} & \phi_{P6_{x}} & \phi_{P6_{P6}} & \phi_{P6_{P7}} \\
\phi_{P6_{x}} & \phi_{P6_{P6}} & \phi_{P6_{P7}} & \phi_{P7_{P6}} \\
\phi_{P6_{P6}} & \phi_{P6_{P7}} & \phi_{P7_{P6}} & \phi_{P7_{P7}} \\
\phi_{P7_{P6}} & \phi_{P7_{P7}} & \phi_{P7_{P7}} & \phi_{P8_{P7}} \\
\phi_{P7_{P7}} & \phi_{P7_{P7}} & \phi_{P7_{P7}} & \phi_{P8_{P7}} \\
\phi_{P8_{P7}} & \phi_{P8_{P7}} & \phi_{P8_{P7}} & \phi_{P9_{P8}} \\
\phi_{P9_{P8}} & \phi_{P9_{P8}} & \phi_{P9_{P8}} & \phi_{P10_{P9}} \\
\phi_{P10_{P9}} & \phi_{P10_{P9}} & \phi_{P10_{P9}} & \phi_{P11_{P10}} \\
\phi_{P11_{P10}} & \phi_{P11_{P10}} & \phi_{P11_{P10}} & \phi_{P11_{P11}}
\end{pmatrix}
\]
Decision Tree Learning Curve Grid

Best Decision Tree Models

SEP Prediction Summary

**Problem:** Predicting Solar Energetic Particles using multivariate time series representing the proton and xray within different energy levels.

**Approach:** Use Vector AutoRegressive (VAR) model to produce features and a decision tree model for the classification.

Conclusion

• We designed a new model for predicting >100 MeV SEP events based on GOES x-ray and proton data.
• We considered the missing data problem by using the same number of SEP and non-SEP coming from GOES-12.
• We found that cross-channels correlation are also an important precursor for SEP occurrence.
• P6_xl_l2 (correlation between P6 and xl channels in 10 minutes) parameter is a strong precursor that appears in both best tree models root nodes.
Thank You 😊