Unsupervised Scalable Representation Learning for Multivariate Time Series
Jean-Yves Franceschi, Aymeric Dieuleveut, Martin Jaggi

To cite this version:

HAL Id: hal-02320167
https://hal.archives-ouvertes.fr/hal-02320167v2
Submitted on 3 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Unsupervised Scalable Representation Learning for Multivariate Time Series

Jean-Yves Franceschi, Aymeric Dieuleveut, Martin Jaggi

1Sorbonne Université, CNRS, LIP6, F-75005 Paris, France (previously at ENS de Lyon and EPFL)
2MLO, EPFL, Lausanne CH-1015, Switzerland
3CMAP, Ecole Polytechnique, Palaiseau, France

Motivation

- Time series are:
  - mostly unlabeled
  - potentially long
  - of unequal length in the same dataset
- Previous work does not tackle these issues simultaneously:
  - most of the time supervised (Bagnall et al. 2017)
  - not scalable (Malhotra et al. 2017)
  - tested on too few datasets with no code available (Malhotra et al., 2017; Wu et al., 2018)
- Objectives of this work:
  - learn unsupervised time series representations,
  - in a scalable way,
  - for time series of potentially unequal lengths,
  - suitable to and extensively tested on various tasks

Encoder Architecture

- We use a neural network based on exponentially dilated convolutions rather than a recurrent network
  - more efficient and parallelizable on modern hardware
  - exponentially increasing receptive field at constant depth
  - good performance on time series for other tasks (Bai et al., 2018, Ismail Feizy et al., 2019)
  - experimentally performs better in our experiments
- We make the network causal
  - maps a sequence to a sequence of the same length
  - each output element only depends on input values with lower time indices
  - can help to save computation time when adding an element to a time series
- The global architecture is sequentially shaped by:
  - a causal network formed with exponentially dilated convolutions associated with:
    - weight normalization
    - leaky ReLU
    - residual connections
    - a global max pooling layer squeezing the temporal dimension and aggregating temporal information in a fixed-size vector
    - a final linear transformation

Training

- Encoder training and testing performed on a single GPU
- No labels used during encoder training
- No hyperparameter optimization
- Open-source code, pretrained models and hyperparameters available
- Examples of dimensionality reduction plots using t-SNE:
Unsupervised Scalable Representation Learning for Multivariate Time Series

Jean-Yves Franceschi,1 Aymeric Dieuleveut,2,3 Martin Jaggi2

1Sorbonne Université, CNRS, LIP6, F-75005 Paris, France (previously at ENS de Lyon and EPFL)
2MLO, EPFL, Lausanne CH-1015, Switzerland
3CMAP, Ecole Polytechnique, Palaiseau, France

Classification

• Protocol:
  • unsupervised training of the encoder on the train dataset
  • training of an SVM with RBF kernel on top of the learned features with the train labels
• Results on the full UCR archive (Dau et al., 2018):
  • we outperform previous unsupervised state-of-the-art methods by a large margin on the few datasets they were tested on
  • we achieve close to state-of-the-art performance when comparing to supervised methods

Tests were also performed on multivariate time series

Additional Features

• Our unsupervised method can be applied in a sparse labeling setting, where it outperforms state-of-the-art deep neural networks
• Learning a one-nearest-neighbor classifier allows to outperform DTW which uses the same classifier on raw data
• The learned representations are transferable across datasets

Moving Average Prediction

• IHEPC dataset:
  • minute-averaged electricity consumption of a single household for four years
  • single unlabeled time series of length \( \approx 2,000,000 \)
  • Encoder on such a long time series is trained in a few hours
  • Linear regressors on raw data versus learned representations for moving average prediction:
    • task: predict next day / quarter average from the previous day / quarter data
    • regressors on raw data show slightly better results
    • regressors on learned representations are much more efficient
  • The learned representations can be leveraged at different time scales

Table: Results obtained on the IHEPC dataset.

<table>
<thead>
<tr>
<th>Task</th>
<th>Metric</th>
<th>Representations</th>
<th>Raw values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>Test MSE</td>
<td>8.92 ( \times 10^{-2} )</td>
<td>8.92 ( \times 10^{-2} )</td>
</tr>
<tr>
<td></td>
<td>Wall time</td>
<td>12s</td>
<td>3min 1s</td>
</tr>
<tr>
<td>Quarter</td>
<td>Test MSE</td>
<td>7.26 ( \times 10^{-2} )</td>
<td>6.26 ( \times 10^{-2} )</td>
</tr>
<tr>
<td></td>
<td>Wall time</td>
<td>9s</td>
<td>1h 40min 15s</td>
</tr>
</tbody>
</table>

References


