

Time series classification using time-frequency analysis and Convolutional Neural Networks

Domen Kavran

Faculty of Electrical Engineering and Computer Science, University of Maribor
Koroška cesta 46, 2000 Maribor, Slovenia
E-mail: domen.kavran@student.um.si

Abstract

Technological advances in various industries have led to an increase in the amount of data in which patterns, trends and useful informations are hidden. Data are quantitatively large, complex and generated at high speed, which makes them impossible to process with traditional methods. A common kind of sequential data are time series, where time order is important. Regression applications are often based on time series, but they can also be used for classification purposes. In recent years, deep learning has proven to be very successful in classifying complex time series data. In this paper, a new method for time series classification is presented using deep learning. Different time-frequency analysis methods are applied on time series in order to obtain time-frequency representations, which serve as input data to convolutional neural networks. By combining various time-frequency representations, the proposed method achieved average classification accuracy of 92,28 %.

1 Introduction

Data collection is being introduced into different areas of industries where this has not been done before. Multitudes of different sensors with high sampling frequencies gather complex data. Such an environment produces large volumes of varied data that are obtained at high velocity. This kind of data is named "big data". Important types of big data are time series, which represent values of a given quantity over time. Physical quantities, such as electric current, temperature, and velocity are observed most commonly [1]. Time series are obtained from different sources, ranging from mobile phones and Internet of Things (IoT) devices to industrial machines and medical equipment [2]. Acquired time series can be noisy and may have information gaps. Sampled time series is treated as a signal consisting of complex sinusoids. These oscillate with different frequencies. Each sinusoid reaches a certain amplitude and has an offset, so-called phase. Sinusoids are present at different time points in a signal. Its amplitude, phase and energy density in time and frequency are analyzed using time-frequency analysis methods, which are presented in the following Sections.

Various methods calculate time-frequency representations, which are limited by the ratio of time and frequency resolution. Time series are, thus, represented as two-dimensional images in the time-frequency plane, where both axes represent time t and frequency f , with signal properties being expressed in grayscale shades. Time-frequency representations are used as input data to convolutional neural networks to perform classification. Most of the time-frequency representation algorithms have long been developed, but they have recently been introduced to the field of Deep Learning as part of the data preparation process [3]. The challenges of time series classification are present in many research areas - from identification of anomalies in financial markets to the automated recognition of brain and cardiovascular diseases [2].

Feature selection is a challenging process, as domain knowledge is often needed. An alternative approach is offered by Convolutional Neural Networks (CNNs), which acquire information-rich features from multidimensional input data using an array of adaptive convolutional filters, called kernels [4]. Many methods for time series classification using CNNs have been proposed throughout the years. Residual neural network (ResNet) is currently considered to be the state-of-the-art deep neural network architecture for time series classification [5]. A recently developed ensemble of five deep CNNs, named InceptionTime, has proven to be very competitive with the state-of-the-art algorithm HIVE-COTE [5].

The proposed CNN architectures use two-dimensional and three-dimensional convolutional kernels. Two-dimensional kernels calculate features in time-frequency space. By combining various time-frequency representations, different local characteristics of the signal are used for three-dimensional feature extraction.

This paper consists of five sections. The next section presents data preparation using Hilbert transform and time-frequency analysis methods. The third section presents CNNs and proposed architectures for time series classification. Classification results are presented in the fourth section. The conclusions are given in the last section.

2 Time series analysis

The analysis of a time series, treated as a signal x , is performed using mutually independent methods and mathematical operations performed in a sequence called a pipeline. Each pipeline consists of several steps. The output of the i -th step is the input to the $i+1$ -th step. The last step is an exception, as its output is the result of the pipeline. The proposed pipeline has the following steps:

1. Calculation of the analytic signal x_a using the Hilbert transform,
2. Calculation of phase or energy density by using one of the following time-frequency analysis methods:
 - Short-Time Fourier Transform,
 - Smoothed Pseudo Wigner-Ville Distribution,
 - Continuous Wavelet Transform,
3. Normalization of values to the interval $[0, 1]$.

Use of an analytic signal makes a frequency spectrum more interpretable, but its use is also mandatory for some time-frequency analysis methods [6]. Phase and energy density are unique signal characteristics in time and frequency. Normalization, as a final step, ensures CNNs learn faster, with input values having equal scales without distortions. Methods in pipeline are described in the following subsections.

2.1 Hilbert transform

Hilbert transform performs phase manipulation without affecting the amplitude spectrum of the signal. The Hilbert transform \mathcal{H} of a signal x is defined by (1), where $h = \frac{1}{\pi t}$ and the Cauchy principal value is denoted by P [6]. The transform is used to calculate an analytic signal that is complex and does not contain negative frequency content. The analytic signal is written by (2) [6].

$$x_H(t) = \mathcal{H}[x(t)] = x(t) * h(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (1)$$

$$x_a(t) = x(t) + ix_H(t) \quad (2)$$

2.2 Short-Time Fourier Transform

Short-Time Fourier Transform (STFT) is a time-frequency analysis method which calculates Fourier transform on segments of a non-stationary signal. Multiplying segments with window function w reduces spectral leakage. Short-Time Fourier Transform has a constant frequency and time resolution that depends on the length of the segments. Computing squared amplitude is represented by a spectrogram, which shows energy density in time and frequency. Short-Time Fourier Transform is defined by (3) [7].

$$STFT_x^w(t, f) = \int_{-\infty}^{\infty} x(\tau) w^*(\tau - t) e^{-i2\pi f\tau} d\tau \quad (3)$$

Short-Time Fourier Transform was demonstrated on a recording of a bird chirp, that was sampled at 22050 Hz. Hann window function was applied on signal segments, which overlapped by half of their length. The

spectrogram of the bird chirp is shown in Figure 1 and the thresholded phase spectrum in time and frequency is shown in Figure 2.

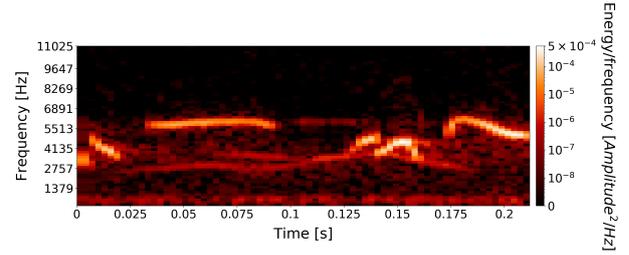


Figure 1: Spectrogram of the bird chirp.

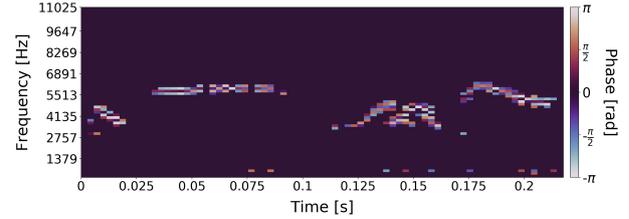


Figure 2: Thresholded phase spectrum in time and frequency of the bird chirp.

2.3 Smoothed Pseudo Wigner-Ville Distribution

Among many methods for time-frequency analysis, Wigner-Ville Distribution is considered the most advanced. It represents energy density, the same as a spectrogram, but with higher resolution, although artifacts called cross terms are present. Smoothed Pseudo Wigner-Ville Distribution (SPWVD) reduces unwanted cross terms with time-frequency filtering; however, time and frequency resolution are deteriorated. Smoothed Pseudo Wigner-Ville Distribution is defined with (4), where $g(s)$ and $h(\tau)$ are window functions for smoothing in time and frequency [8]. The result of Smoothed Pseudo Wigner-Ville Distribution on a recording of a bird chirp is shown in Figure 3.

$$SPWVD_x(t, f) = \int_{-\infty}^{\infty} h(\tau) \left[\int_{-\infty}^{\infty} g(s - t) x(s + \frac{\tau}{2}) x^*(s - \frac{\tau}{2}) ds \right] e^{-i2\pi f\tau} d\tau \quad (4)$$

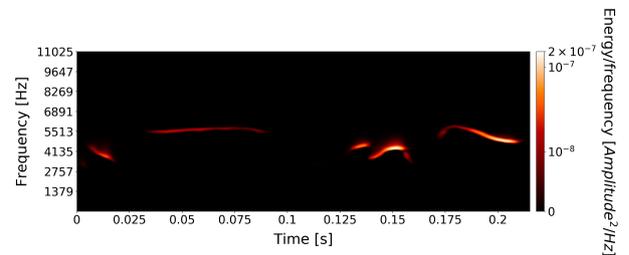


Figure 3: SPWVD of the bird chirp.

2.4 Continuous Wavelet Transform

Continuous Wavelet Transform (CWT) is an alternative to Short-Time Fourier Transform. An important property of Continuous Wavelet Transform is that it does not use Fourier transform. Instead, convolution of the signal is

performed using a function, called a wavelet. Time localization is carried out with translation τ of mother wavelet ψ . Its length is adjusted with scale s , thus effecting time and frequency resolution. Computing squared amplitude is represented by a scalogram, which shows energy density in time and scale. Continuous Wavelet Transform is defined by (5) [9]. Results of Continuous Wavelet Transform with a complex Morlet wavelet of a bird chirp are visible on Figures 4 and 5.

$$CWT_x^\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-\tau}{s}\right) dt \quad (5)$$

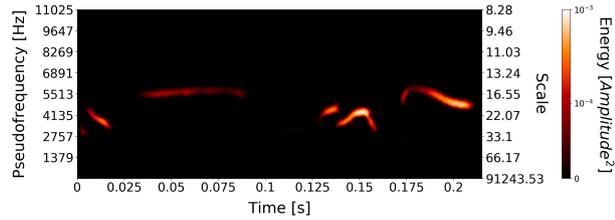


Figure 4: Scalogram of the bird chirp.

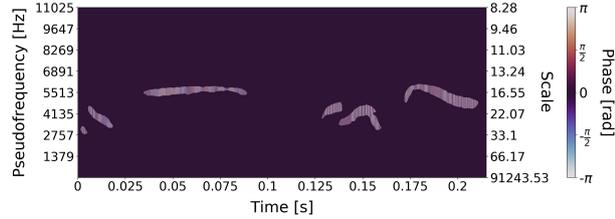


Figure 5: Thresholded phase spectrum in time and scale of the bird chirp.

3 Convolutional Neural Networks

The most established algorithms in the field of Deep Learning are CNNs, which are designed to learn spatial hierarchies of features automatically by using convolutional layers, pooling layers and fully connected layers. They are often used to solve computer vision problems in industrial processes and various fields of medicine. CNNs achieve exceptional results at segmentation and recognition of objects in images and at processing natural language [4].

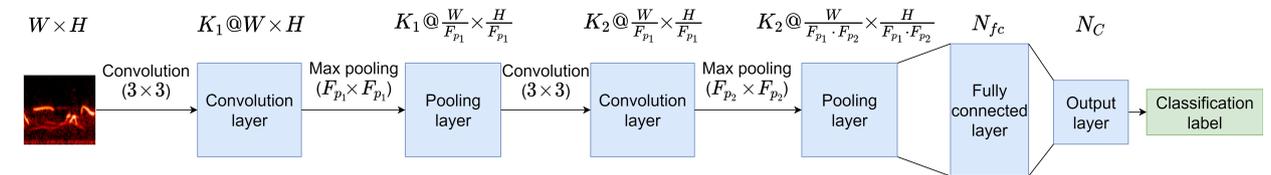


Figure 6: Architecture CNN-I for time series classification, using individual time-frequency representations as the input.

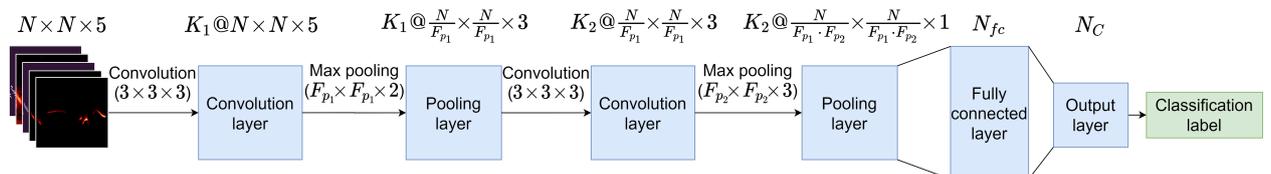


Figure 7: Architecture CNN-C for time series classification, using combined time-frequency representations as the input.

3.1 Proposed CNN architectures

Figures 6 and 7 show the proposed CNN architectures for classification of time series of length N . The CNN-I architecture is designed for time-series classification, based on individual time-frequency representations. The CNN-C architecture performs time series classification with combined time-frequency representations, used as input. The input sizes $W \times H$ represent the width and height of the time-frequency representations, and the input sizes $N \times N \times 5$ are intended for combined time-frequency representations. The hyperparameter K represents the number of convolutional kernels, and F_p the size of pooling filters. Convolution layers and pooling layers use zero padding. The number of neurons in a fully connected layer is defined by the hyperparameter N_{fc} . The number of neurons in the output layer is equal to the number of classification classes N_C [4].

4 Results

Classification was performed on the datasets with pre-segmented time series described in Table 1 [10, 11]. All datasets (except Epileptic Seizure Recognition) had a predetermined train and test set. For individual time series dataset, five pipelines were defined to calculate the following normalized outputs for each time series of length N :

- Phase spectrum as a result of STFT,
- Spectrogram,
- Phase spectrum as a result of CWT,
- Scalogram,
- Smoothed Pseudo Wigner-Ville Distribution.

The output of each pipeline is a time-frequency representation of the time series, which is then used as input data to one of five CNNs having CNN-I architecture. Because the sizes of representations are not equal, they need to be scaled to size $N \times N$ with nearest neighbors interpolation before combining them and forwarding them to a CNN having CNN-C architecture. The hyperparameters of each CNN were adjusted based on 10-fold cross validation.

Table 1: Time series datasets.

Dataset name	Description	Time series length	No. of classes
ECG5000	ECG recordings of 15 patients with five categorizations of cardiovascular diseases	140	5
InsectEPGRegularTrain	EPG signals of insect interaction with plants	601	3
Epileptic Seizure Recognition	Sections of EEG recordings of 500 individuals with epilepsy attacks	178	5
MelbournePedestrian	Hourly number of pedestrians at ten locations in the city of Melbourne	24	10
ElectricDevices	Electricity consumption of seven groups of appliances in UK households	96	7
Wafer	Control measurements of sensors in the processing of silicon wafers	152	2
PowerCons	Electricity consumption of households in summer and winter	144	2

The results of the proposed classification method are shown in Table 2. Each column contains classification accuracies, obtained with selected time-frequency representations, and used as inputs to CNNs with a chosen architecture (CNN-I or CNN-C). The highest classification accuracy of each dataset is marked with bold text. Using individual time-frequency representations as inputs to CNNs with architecture CNN-I, the highest classification accuracies in all datasets were

achieved with phase spectrums or scalograms, obtained with Continuous Wavelet Transform.

Convolutional Neural Networks with architecture CNN-C, that performed classification based on combined time-frequency representations, achieved higher classification accuracies of datasets ECG5000, InsectEPGRegularTrain, Epileptic Seizure Recognition and MelbournePedestrian.

Table 2: Time series classification accuracies.

Dataset name	STFT - phase spectrum (CNN-I)	Spectrogram (CNN-I)	CWT - phase spectrum (CNN-I)	Scalogram (CNN-I)	SPWVD (CNN-I)	Combined representations (CNN-C)
ECG5000	93,29%	92,98%	93,69%	93,51%	93,33%	94,27%
InsectEPGRegularTrain	95,58%	83,13%	99,60%	99,60%	83,13%	100,00%
Epileptic Seizure Recognition	91,04%	92,35%	94,87%	96,70%	95,78%	97,39%
MelbournePedestrian	61,96%	63,56%	83,22%	75,30%	68,36%	88,39%
ElectricDevices	57,23%	65,27%	61,47%	69,14%	63,91%	68,28%
Wafer	98,96%	99,04%	99,43%	99,53%	98,60%	99,33%
PowerCons	91,11%	98,89%	96,11%	98,89%	96,67%	98,33%

5 Conclusion

In this paper, a new method for time series classification was presented, where data preparation pipelines compute time-frequency representations, which are then used as input to Convolutional Neural Networks for classification. Convolutional Neural Networks were most successful at classifying phase spectrums and scalograms, which were calculated with Continuous Wavelet Transform, achieving an average accuracy of 90,07 %. Classification accuracies of some time series datasets have been improved by combining different time-frequency representations, achieving an average accuracy of 92,28 %. In the future, the proposed Convolutional Neural Network architectures will be adapted to perform multivariate time series classification.

Acknowledgment

I thank Assist. Prof. Dr. Niko Lukač for guidance and professional assistance in writing my Master's thesis, which was the basis for this paper.

References

- [1] N. Elgendy and A. Elragal. Big Data Analytics: A Literature Review Paper. *Advances in Data Mining: Applications and Theoretical Aspects*. 14th Industrial Conference on Data Mining, St. Petersburg, 16.–20. July 2014. Leipzig: Springer, 214–227, 2014.
- [2] G. A. Susto, A. Cenedese and M. Terzi. Time-Series Classification Methods: Review and Applications to Power Systems Data. *Big Data Application in Power Systems*. Amsterdam: Elsevier, 179–220, 2018.
- [3] J. Zhang, J. Tian, Y. Cao, Y. Yang and X. Xu. Deep time–frequency representation and progressive decision fusion for ECG classification. *Knowledge-Based Systems*, 190, 2020.
- [4] R. Yamashita, M. Nishio, R. K. G. Do and K. Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4):611–629, 2018.
- [5] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, et al. InceptionTime: Finding AlexNet for Time Series Classification, 2019.
- [6] M. Johansson. *The Hilbert Transform*. Växjö: Växjö University, 2005.
- [7] F. Hlawatsch and F. Auger. *Time-Frequency Analysis: Concepts and Methods*. London: ISTE, 2008.
- [8] A. Djebbari and F. Bereksi-Reguig. Detection of the valvular split within the second heart sound using the reassigned smoothed pseudo Wigner–Ville distribution. *BioMedical Engineering OnLine*, 12(1):1–21, 2013.
- [9] L. Aguiar-Conraria and M. J. Soares. The Continuous Wavelet Transform: moving beyond uni and bivariate analysis. *Journal of Economic Surveys*, 28(2):344–375, 2014.
- [10] A. Bagnall, E. Keogh, J. Lines, A. Bostrom and J. Large. UEA & UCR Time Series Classification Repository, 2020.
- [11] D. Dua and C. Graff. UCI Machine Learning Repository. Irvine: University of California, School of Information and Computer Science, 2020.