Neural network simulating the Schuster periodogram

Vladislav V. Topinskiy and Roman V. Baluev

Abstract. The search for periodic components in a time series is an important aspect of data analysis. In most cases, Schuster's periodograms or Lomb-Scargle periodograms are used depending on the homogeneity of the distribution of the original data over time. Calculating spectra is not a computationally intensive task; however, difficulties arise when processing large quantities of time series data and assessing the existence of periodic components within them. For preliminary analysis of large data sets, a convolutional neural network simulating the operation of Schuster's periodogram is suitable.

Introduction

This work represents our initial step towards accelerating the primary processing of signals in the task of exoplanet detection. Initially, a two-layer perceptron was designed to determine the existence of a sinusoidal component in a signal consisting of 128 samples by means of its Fourier transform. Each layer is defined by the following formula:

$$x_k^{i+1} = f\left(\left(\sum_{j=1}^{n} w_j^i x_j^i\right) + b_k^i\right), \tag{1}$$

where f – layer's activation function (sigmoid were used), w_j^i – weight, b_k^i – bias. Training time series were generated for network training and subsequent testing. Bayes factor, as described in the paper [1], was used to assess prediction accuracy.

The first attempt was not very successful: the accuracy reached only 90% for series with high amplitudes. Increasing accuracy required increasing the number of neurons, which in turn resulted in numerous "extra" connections in the network that needed to be optimized through training, consequently increasing the amount of required data. Therefore, it was decided to restructure the network to reduce the number of neural connections while improving accuracy. In this case, the best alternative was to introduce convolutional layers into the structure.

Model description

Replacing regular layers with convolutional layers reduces the number of trainable connections and allows for an increase in the number of neurons in each layer. To

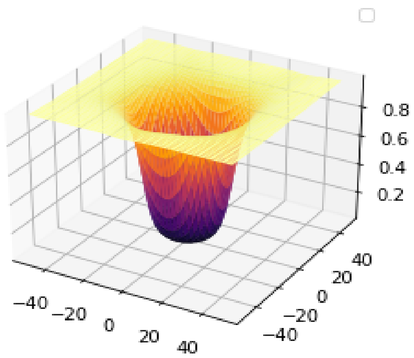

FIGURE 1. Network structure, first layer. Response graph of a 16-neuron block to a pair of real and imaginary parts of a signal sample

optimize training, weights from pre-trained models' layers were used. The network structure now consists of 2 convolutional layers processing a pair of Fourier components (Figure 1.), a technical flattening layer, and 2 regular layers responsible for finding the maximum and outputting the result as the probability of a sinusoidal component in the signal. The neural network was implemented using Python 3.8, and the network structures were taken from the module keras. Training took place over 300 epochs, with the adadelta optimizer, accuracy metric, and binary cross-entropy loss function. The choice of optimizer was based on its precise and rapid weight minimization, as determined through empirical testing.

Tests

Tests were conducted on synthesized datasets with approximately $N \sim 10^6 - 10^7$. The amount of data containing a sinusoidal component and data consisting solely of noise was equal. This volume allowed achieving a signal detection accuracy of 99% of the theoretical maximum with a small number of trainable network neurons (Figure 2.). Additionally, no overfitting issue affecting the network's response was observed.

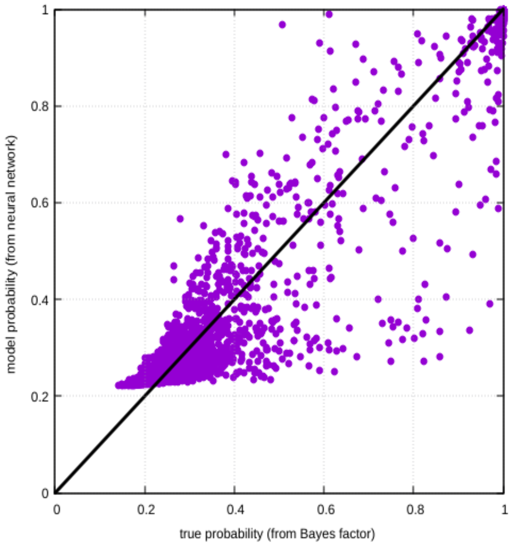

FIGURE 2. The ratio of the model probability of the existence of a signal with the theoretically possible

Conclusion

This work presented a brief description of a convolutional neural network model solving the signal detection problem. The described structure is currently not well-suited for real data; hence work is underway to expand its functionality, specifically introducing weights to time series and processing non-uniform series (simulating the operation of Lomb-Scargle periodograms). The synthesis of training datasets will also be revised for more efficient training. These steps will enable obtaining results from real data.

References

[1] Roman V. Baluev, *Comparing the frequentist and Bayesian periodic signal detection: rates of statistical mistakes and sensitivity to priors.* preprint (2022), available at `https://arxiv.org/abs/2203.08476`.

Vladislav V. Topinskiy
Saint Petersburg State University
Saint Petersburg, Russia
e-mail: `st076660@student.ru`

Roman V. Baluev
Saint Petersburg State University, 7–9 Universitetskaya Emb., St Petersburg 199034, Russia
Special Astrophysical Observatory, Russian Academy of Sciences, Nizhnii Arkhyz, 369167, Russia
e-mail: `r.baluev@spbu.ru`