Neural network simulating the Lomb-Scargle periodogram

Vladislav V. Topinskiy and Roman V. Baluev

Abstract. A convolutional neural network is developed for detecting a sinusoidal signal in unevenly sampled time series with periodic gaps and Gaussian white noise. The model, trained with binary cross-entropy, approximates the Bayesian posterior probability in the limit of infinite sample size. On test data, the network's root-mean-square deviation from the theoretical detection probability is about 0.07, while its processing speed exceeds numerical Bayesian modelling by 5–6 orders of magnitude. A critical sensitivity to initial weights is revealed, and empirical distributions for their initialization are proposed.

Introduction

The search for periodic signals is an important problem in astronomy (exoplanet detection, variable stars) and other fields. This work investigates the potential of artificial neural networks for detecting a sinusoidal signal in noise. The main goal is to assess the capabilities of neural networks, identify "pitfalls", and outline directions for further development. As a test case, we consider the detection of a harmonic signal in an uneven time series with gaps and Gaussian white noise. The neural network effectively emulates the Lomb–Scargle periodogram [1], [2].

Bayesian interpretation of the neural network classifier

When a neural network is trained with the binary cross-entropy (BCE) loss function, minimisation in the limit of an infinite sample leads to the network output approximating the posterior probability of signal presence, $g_1(x) = P(\omega_1|x)$. This generalises the known result for the L_2 mere [3]. The derivation is based on variational calculus: the BCE functional attains its global minimum at $F^*(x) = g_1(x)$. Second variations show that BCE is more sensitive to errors in regions where the posterior probability is close to 0 or 1, i.e. near the boundaries.

Uneven time series with gaps

The time series may be uneven (a model with periodic gaps; the parameters λ and μ are chosen randomly: $\lambda \sim \mathcal{B}(1,2)$ and $\mu \sim \mathcal{B}(1,1)$), and may contain an arbitrary offset and noise variance. To remove the influence of offset and variance, each series is centred and normalised. The input data are supplemented with information about the observation times through the spectral window. For each frequency f_k, both $\Omega(f_k)$ and $\Omega(2f_k)$ are included in the feature vector (the latter is related to the generalised Lomb–Scargle periodogram [4]). A switch to a convolutional architecture is made: the first layer is a convolutional kernel of size $1 \times 6 \times m$ (where 6 is the number of input channels: real and imaginary parts of the DFT, $\Omega(f_k)$, $\Omega(2f_k)$, the second layer is a convolution $1 \times m \times 1$, followed by a fully connected layer and the output. This reduces computational cost and allows scaling.

The problem of initial weights

Training with random weights practically never converged to a good minimum due to "dead" neurons and subtraction of nearly equal numbers. An empirical initialisation for the weights of the first layer was proposed, based on modelling parameter distributions to avoid pathologies. For each frequency block, the quantities $H_i = \sqrt{(C_i^2 + S_i^2)}$, $R_i = -B_i/H_i$, $\alpha_i = atan2(S_i, C_i)$ were introduced, where C_i, S_i are the weights for the real and imaginary parts of the DFT, and B_i is the bias. It was found that H_i and R_i are close to normal, α_i to uniform, and the output layer weights D_i to half-normal. Additionally, it was empirically discovered that the weights handling $\Omega(2f)$ must be related to the DFT weights by $\frac{S^{\text{“}}}{C^{\text{“}}} = \frac{SC}{C^{\text{“}2} - S^{\text{“}2}}$, $S^{\text{“}2} + C^{\text{“}2} \approx 0.5(S^2 + C^2)$. Setting such initial approximations significantly improved convergence. The weights for $\Omega(f)$ turned out to be not very informative and had little effect on accuracy. This starting point allowed reaching a meaningful minimum. The optimal block size is 16 or 32 neurons.

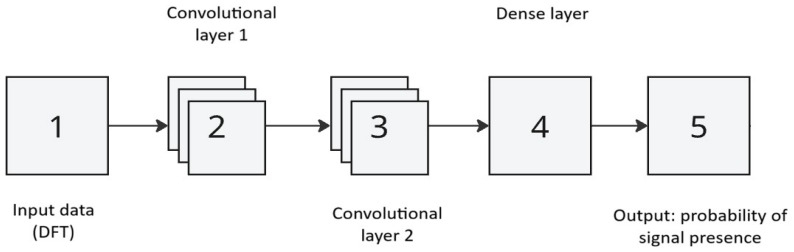


FIGURE 1. Neural network structure.

Results

For a test sample of 40,000 series, the Bayesian posterior detection probability of 0.3851. The convolutional network with $m = 16$ or 32 achieved BCE of about 0.4048 and 0.4047, respectively, and a root-mean-square error of about 0.07. A comparison on a two-dimensional histogram showed good agreement with the theoretical posterior probability, with the majority of points concentrated in the corner regions (probabilities near 0 or 1), where BCE is most sensitive [5].

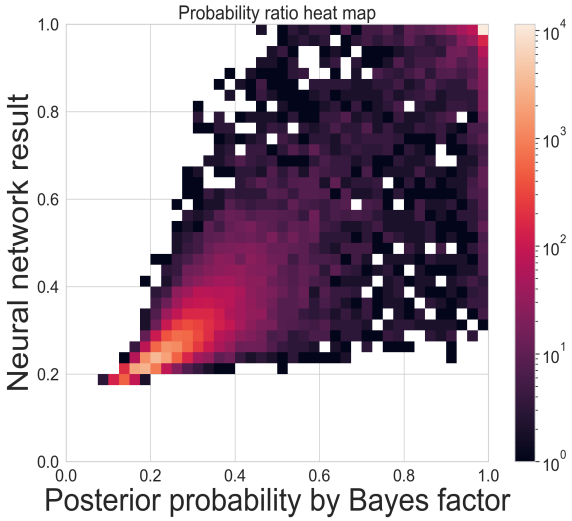


FIGURE 2. Comparison of the final neural network model output with the Bayesian posterior probability of signal presence. A two-dimensional histogram is shown, where colour indicates the density of test points (40,000 points in total from the validation sample).

Discussion and prospects

The main advantage of the neural network model is its processing speed: running through 10^6 series takes about 1 minute on a GPU, while computing the Bayesian probability for 40,000 series on a CPU took about a week (a gap of 5–6 orders of magnitude). This makes the method suitable for processing large astronomical catalogues. However, training is extremely sensitive to initial weights – without an empirical "hint", convergence to the global minimum is unlikely. Future plans include:

- increasing the number of observations K and the spectral range;
- adapting the model for non-sinusoidal signals (e.g., highly eccentric orbits);
- accounting for correlated noise (non-uniform spectrum).

Thus, the work demonstrates that neural networks can effectively approximate Bayesian inference in periodic signal detection, but they require careful tuning of initial conditions. The proposed approach can serve as a foundation for more complex applications.

References

[1] Lomb N.R. *Least-squares frequency analysis of unequally spaced data Astrophysics and Space* Science 39 (2), 447–462 (1976). doi:10.1007/BF00648343/

[2] Scargle J. D. *Studies in astronomical time series analysis. II- Statistical aspects of spectral analysis of unevenly spaced data* The Astrophysical Journal 263, 835–853 (1982). doi:10.1086/160554/

[3] Ruck D. W., Rogers S. K., Kabrisky M., Oxley M. E., Suter B. W. *The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function* IEEE Transactions Neural Networks 1 (4), 296–298 (1990). doi:10.1109/72.80266

[4] Zechmeister M., Kurster M. *Generalized Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms* Astronomy and Astrophysics 496 (2), 577–584 (2009). doi:10.1051/0004-6361:200811296

[5] Vladislav V. Topinskiy, Roman V. Baluev, *Application of artificial neural networks in the problem of periodic signal detection* Vestnik St.Petersb. Univ.Math. 59, 234–243 (2026). `https://doi.org/10.1134/S1063454126700135`.

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`