Challenges in stochastic time series prediction

Maximilian Hornung, Jens Settelmeier

April 22, 2019

Abstract

Predicting the future development of time series is of interest in different areas of computational biology. The time series in biological models exhibit challenging characteristics such as chaotic or stochastic behavior. In this report, time series prediction is done on different kind of biological time series using deep neural networks. In our evaluation, we identify challenges and limitations of this approach, and compare different architectures of deep neural networks with regard to their performance.

1 Introduction

In various kind of diseases, it is important to choose the correct treatment at the current time point. Since the human body is too complex to be described directly using mathematics, special features are modeled with mathematical systems such as the Mackey-Glass equations for respiratory and hematopoietic diseases [MG77]. These models make use of available data about previous information in order to reduce the uncertainty under which a decision is done. This is especially important when the consequences of a decision are severe, for example the decision of breast cancer treatment after radiation. This particular problem can be addressed by using a stochastic differential equation, as done by Oroji *et al* [OOY16].

In different kinds of biological models, chaotic behavior is observed. Even the Mackey-Glass equations obtain chaotic behavior for certain choices of their parameters, as shown by Fischer [Far82]. That means that even small perturbations of the time series prediction increase exponentially, which means that the time series prediction should be as robust as possible against noise.

This gets imporant in particular because another characterstic of dealing with biological data is the presence of noise. In this report, we show that different kinds of noise assumptions can impact the results of time series prediction. We evaluate all our investigated problems with regard to the popular *i.i.d.* random noise as well as memorizing random noise. Since mathematical models of biological systems can not describe their characteristics perfect and error-free, it is assumed that the latter type of noise is more realistic. But unfortunately this kind of noise increases the difficulty of time series prediction, even for simple mathematical functions, as seen in section 3.1.

The remaining report is structured as follows. First, we perform proper time series prediction on the *sine* function and show how different noise models impact the prediction ability of our network models. We conduct an evaluation how these results can be extended to the prediction of ordinary differential equations (**ODE**s) at the example of the differential equations of the harmonic oscillator. After that, we show how the stochasticity of noise impactss the possibility to predict deterministic chaotic time series at the example of the *Mackey Glass* time series. Last but not least, we report how the poor results of numerical approximations on stocastic differential equations (**SDE**s) can be explained based on our previous results.

2 Methods

In our analysis, we use four different architectures and neural networks and optimize their hyperparameters for the respective use case. Since Hornik *et al.* have shown that feedforward neural networks are universal function approximators [HSW89], we evaluate this architecture with varying number of hidden nodes.

After that, we investigate the strength of improvement of using a recurrent neural network. This type of neural network has already been applied to time series prediction [CMA94], but suffers from the vanishing gradient problem when capturing long-term dependencies in a sequence. Because of that, we decide to evaluate only the Long Short-Term Memory (**LSTM**) network architecture [HS97], which was successfully applied in various time series prediction tasks like anomaly detection [MVSA15], stock price [FK18] and protein disorder prediction [HYPZ16].

In the last years, convolutional neural networks (**CNN**s) have improved the results in image classification [KSH12] and other computer vision tasks. A **CNN** learns features from the data in a hierarchical way, for example combining pixels to edges, edges to more complex forms etc. until a high-level classification can be done. The large success in computer vision has inspired researchears in time series prediction to also apply **CNN**s [CCC16, BBO17], so we also evaluate this architecture in our analysis.

Last but not least, we analyse the **N**onlinear **A**uto**R** egressive models with e**X**ogenous input (**NARX**). These networks are a recurrent neural networks that relinquish of feedback from the hidden state and only use feedback from the output state. It has been shown by Siegelmann *et al.* that even with the limited feedback **NARX** has computing capabilities of a Turing machine [SHG97] and is therefore a universal function approximator, too. This type of neural network was not only used for chaotic time series prediction [Dia08], but also provides current state of the art performance in time series prediction as a attention-based neural network [QSC$^+$17].

All implementation is done using the `Python` programming language in version 3.6.7. The time series data is created and loaded in the `numpy` framework in version 1.16.1. In order to train the neural networks both fast and elegant, we use the `keras` (version 2.2.4) with the `tensorflow` backend in version 1.13.1.

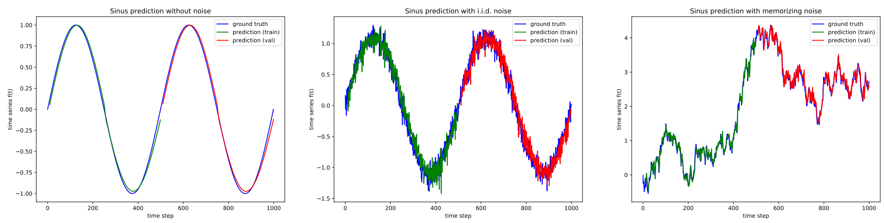

Figure 1: Impact of different kinds of noise on the time series prediction using a feedforward neural network with one hidden layer. Number of hidden units stays 10 in all simulations. Gaussian distributed noise with $\sigma = 0.1$.

The reason for this choice is the tight integration between `keras` and `numpy` that simplifies and increases the speed of our software development. Before running the experiments, the random number generator of `numpy` is set to the seed 0 to ensure reproducibility of our results.

3 Evaluation

3.1 Time series prediction of periodic functions

In the first part of the analysis, we analyze the capability of different neural network architectures to perform time series prediction on the periodic *sine* function. This task can be considered simple, because the different parts of the sine wave occur multiple times in the data. It is therefore interesting how robust the network architectures are with regard to different kinds of noise.

In this section, we provide information about the last 10 time sequence points to the network and want to predict the next point.

In Figure 1, we see that a densely connected neural network with one hidden layer of 10 nodes is capable of performing the time series prediction. Even if we apply *i.i.d.* gaussian noise with standard deviation $\sigma = 0.1$, the network can still be trained to fit the data correctly. But if the noise is not applied independently in each timestep, the convergence takes more time steps as pointed out in Figure 2. This effect is especially strong for the **LSTM** architecture, where even 100 epochs are not enough for the validation loss to converge, compared to the other approaches that need only 20 epochs.

Since the noise we apply on the data has a standard deviation $\sigma = 0.1$, it is obvious that a perfect time series prediction is not be possible. We expect that the minimum possible validation loss is as high as the variance (σ^2), and can reproduce that in all neural network architectures as shown in Table **??**.

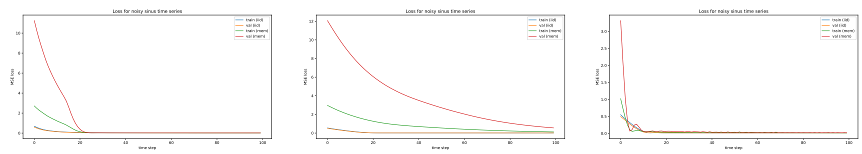

Figure 2: MSE loss over the number of trained epochs on training and on validation dataset. On the left, the classical feedforward network was used. The middle plot shows the performance of the LSTM and the right one plots the loss values of the convolutional network.

3.2 Time series prediction of ODEs

3.3 Mackey Glass time series prediction

In order to model diseases related to dynamic respiratory and hematopoietic diseases, Mackey *et al.* proposed the mackey-glass equations, a kind of first-order nonlinear differential delay equations [MG77]. If the delayed time ($x_\tau = x(t-\tau)$) exceeds the delay $\tau > 16.8$, then equation 1 behaves chaotic [Far82].

$$\frac{dx}{dt} = \beta \cdot \frac{x_\tau}{1 + x_\tau^n} \tag{1}$$

The first approach to predict the short-time behavior of chaotic time series was done by Farmer *et al.* who proposed a `local approximation` technique [FS87]. After improval of predictions using support vector machines by Mller *et al.* [MSR+97], the focus in research shifted towards artifical neural networks which enable even better predictions. Two of the latest developments are the usage of Wavelet Networks [AZ13] and particle swarm optimization [LCSL+16].

3.4 SDE time series prediction

4 Conclusion

References

[AZ13] Antonios K Alexandridis and Achilleas D Zapranis. Wavelet neural networks: A practical guide. *Neural Networks*, 42:1–27, 2013.

[BBO17] Anastasia Borovykh, Sander Bohte, and Cornelis W Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*, 2017.

[CCC16] Zhicheng Cui, Wenlin Chen, and Yixin Chen. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*, 2016.

[CMA94] Jerome T Connor, R Douglas Martin, and Les E Atlas. Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks*, 5(2):240–254, 1994.

[Dia08] Eugen Diaconescu. The use of narx neural networks to predict chaotic time series. *Wseas Transactions on computer research*, 3(3):182–191, 2008.

[Far82] J Doyne Farmer. Chaotic attractors of an infinite-dimensional dynamical system. *Physica D: Nonlinear Phenomena*, 4(3):366–393, 1982.

[FK18] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669, 2018.

[FS87] J Doyne Farmer and John J Sidorowich. Predicting chaotic time series. *Physical review letters*, 59(8):845, 1987.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[HYPZ16] Jack Hanson, Yuedong Yang, Kuldip Paliwal, and Yaoqi Zhou. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics*, 33(5):685–692, 2016.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[LCSL+16] CH López-Caraballo, I Salfate, JA Lazzús, P Rojas, M Rivera, and L Palma-Chilla. Mackey-glass noisy chaotic time series prediction by a swarm-optimized neural network. In *Journal of Physics: Conference Series*, volume 720, page 012002. IOP Publishing, 2016.

[MG77] Michael C Mackey and Leon Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300):287–289, 1977.

[MSR+97] K-R Müller, Alexander J Smola, Gunnar Rätsch, Bernhard Schölkopf, Jens Kohlmorgen, and Vladimir Vapnik. Predicting time series with support vector machines. In *International Conference on Artificial Neural Networks*, pages 999–1004. Springer, 1997.

[MVSA15] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings*, page 89. Presses universitaires de Louvain, 2015.

[OOY16] Amin Oroji, Mohd Omar, and Shantia Yarahmadian. An ito stochastic differential equations model for the dynamics of the mcf-7 breast cancer cell line treated by radiotherapy. *Journal of theoretical biology*, 407:128–137, 2016.

[QSC+17] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017.

[SHG97] Hava T Siegelmann, Bill G Horne, and C Lee Giles. Computational capabilities of recurrent narx neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 27(2):208–215, 1997.