

Towards an Open Book Architecture for Deep Learning Networks: Data Properties and Network Architectures - Evidence from Time Series Analytics

Vera Fleißner¹, Kai Heinrich¹, Michael Seifert¹

¹ TU-Dresden, Germany

Abstract. In the field of time series predictions there are effective and consistent inaccuracies which potentially cause wrong economic decisions. Neural networks are seen as a potential option to improve predictions and are already used in some cases. Nevertheless it is unclear so far for which kind of time series data the use of neural networks is the right choice but also what kind of neural networks are feasible for which data characteristics. The main goal of this research-in-progress paper is therefore to conduct a between different time series characteristics and network architectures by using a descriptive analysis based on an extensive literature review. This goal is a step towards developing design guidelines for (deep) neural networks in the context of time series analysis.

Keywords: time series, deep learning, neural network, forecasting, prediction, survey.

1 Introduction

The field of time series prediction through intelligent systems using Artificial Neural Networks (ANN) is steadily growing [1]. A reason for that lies within the potential of complex intelligent systems such as Long Term Short Memory (LSTM) or Convolutional Neural Networks (CNN) to overcome shortcomings of traditional algorithms such as the vast amount of statistical assumptions to model complex non-linear time series [2].

Neural Networks are especially popular in the field of Big Data Analytics and its subsequent domains like Natural Language Processing (NLP) and Pattern Recognition, often times ranking first in prediction quality for several benchmark datasets [3].

When it comes to the task of time series prediction, neural networks have widely been adopted for the task [4]. However, time series can exhibit multiple characteristics like non-linearity, noise or deterministic chaos that makes them hard to predict and demands suitable models for certain characteristics to be caught. While time series prediction is applied in many fields such as credit scoring, inflation analysis or stock market prediction, it remains unclear (1) how those black box models catch certain inherent characteristics of time series and (2) how they react to different problem and sample characteristics like sample size or prediction window and (3) how they perform in terms of prediction quality compared to other algorithms [5].

Our main goal within the full research project is to present guidelines on deep learning architectures based on time series data characteristics. Towards this goal we first aim to reveal the research gap of intransparent behavior of deep learning algorithms in the context of time series analysis. We therefore propose the following questions:

- **RQ1:** What time series characteristics are usually present in the data when using neural network architectures?
- **RQ2:** How do different architectures handle different characteristics?

Our paper is structured as follows: We first give some preliminaries on time series characteristics in Section 2. After we described our research methodology in Section 3, we present the results of our descriptive analysis in Section 4 in order to answer research questions RQ1 and RQ2 in Sections 4.1 and 4.2 respectively. The paper closes with a discussion and an outlook in Section 5.

2 Time Series Characteristics

The most common characteristic in a time series is *stationarity*. We can assume a time series as stationary when it has constant mean, variance and covariance for a given time step k regardless of time t . When decomposing time series, we can yield a *trend* and a *seasonality* component, if such effects are present in the current time series. If a time series shows trends and seasonality they are not stationary, since they do not have constant moments throughout the observation interval. Furthermore a time series can show signs of *non-linearity*. Non-linearity can lead to small influences of big shocks and vice versa. Influences in non-linear time series will also depend on the sign of the shock effect [6]. Most of the time series data will show evidence of multiple characteristics [7]. In addition to that, those characteristics might determine each other. In addition to that, the prediction error grows with every additional prediction made upon a previous value, since those values depend themselves on predictions rather than a real data point. Further characteristics include *anomalies* and deterministic *chaos* behavior. Anomalies can occur in several forms: Additive outlier effects one observation that lies within normal range but does not reflect a correct value, might be caused by measure errors; Innovation outlier effects several values (e.g., driving noise during an erroneous interval); changing trend caused by a shock event in the data or a permanent level shift of the measurement system (PLS). Dismissing anomalies can lead to significantly worse and biased predictions [6]. Deterministic chaos as an additional source for seemingly unpredictable randomness, while limiting long term prediction, can be facilitated in short-term prediction models if detected correctly [8]. Time series can also exhibit signs of no deterministic stochastic process at all, which is evidence for the presence of a noisy component. Noise, like many other characteristics is not an exclusive property and can be mixed with deterministic parts, so that a time series can just show signs of partial noise [9].

3 Research methodology

As a basis for our descriptive analysis we employ a systematic literature review [10, 11]. We derive constraints for the review from our RQ1 and RQ2. The *technical constraints* of our literature review can be summarized as follows: Only peer-reviewed journals were included; English is the written language of the paper; Papers are accessible; List of databases: IEEE; Science Direct; AIS; Springer; JSTOR; Emerald; ACM; Web of Science. The *content constraints* can be derived by the research questions: The focus of the literature review was on application of approaches to time series analysis, so that we excluded papers that were solely based on theoretical algorithm introductions. We combined a group of search terms from three different topic areas: algorithm, task and domain. The task and domain are given by “prediction” and “time series” respectively. For the area of algorithm we compiled a list of commonly known methods applied within the field of intelligent systems (e.g., neural networks, deep learning). The summary of the search terms that were yielding at least one result can be found in Table 1.

Table 1. Search terms of the literature review

Search Term	# of results
net* AND time series AND fore*	113
net* AND time series AND prediction	93
perceptron AND time series AND fore*	5
perceptron AND time series AND prediction	4
rbf AND time series AND prediction	5
Sum of papers initially found	226

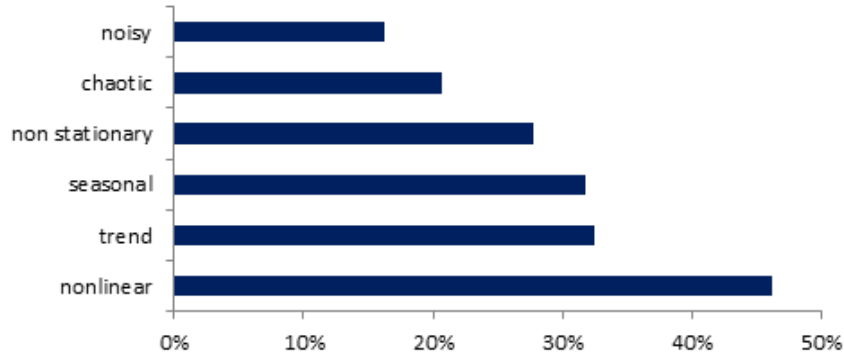
We initially found 226 papers. After we applied the technical constraints and removed duplicates 192 papers were left. We then filtered out papers with unsuitable non numeric time series data (170 papers left) and removed papers that were not topic relevant (e.g., no focus on intelligent systems) to our cause (135). In the next section we conduct a descriptive analysis and first answering RQ1 in Section 4.1 and subsequently RQ2 in Section 4.2

4 Descriptive analysis and results

4.1 Time Series Characteristics

The majority of the found papers is published in the field of computer science. The research goals of the papers can be divided into model comparison, prediction (new problem), improve prediction (existing problem) and improve model. The improve prediction task makes up the majority of the papers (51%), followed by the improve model task (22%). The distribution of time series characteristics within the literature review database is depicted in Fig. 1. We can see that non-linear time series are the subject of a majority of papers along with trend and seasonal components.

Fig. 1. Time Series Characteristics from Literature Review



The different dataset sizes and their frequency of occurrence have a peak at around 500 data points, the majority of the studies does not extend into big data territory in terms of volume. Only a small fraction of the studies use datasets with 10000 observations or more. This is particularly interesting when applying deep learning algorithms that often times long for rather large training datasets. Considering the used algorithms in the studies observed we used a taxonomy approach by [12] to group them. First we classified them by **basic network architecture** where we divide them up into classic ANNs with standard activation functions (denoted “ANN”), ANNs with different functions and different architectures (denoted “no ANN”), e.g., Ridge Polynomial or Abductive Networks and hybrid methods consisting of at least a two stage approach applying classic ANN and other methods, e.g., fuzzy functions (denoted “*hybrid method*”). Secondly we divide them up due to their **properties**, where we distinguished between modular networks (“*modular*”) that can easily be expanded using additional layers, *feedforward* and *feedbackward* architectures, *interlayer* networks that allow connections within each layer, convolutional networks and network architectures that allow recurrence using *time* features. Since hybrid methods can both be related to architecture as well as features (e.g., using hybrid activation functions), we added hybrid to the distinction of properties as well. Fig. 2 shows the correlation between the two distinction schemes (**basic network architecture** as blue, red and green bars vs. **properties** on the x-axis). It can be derived from Fig 2. that some neural network properties occur more often than others. While the property of *modular* networks is uniformly distributed, networks that do not have a classical ANN as base network (*no ANN*) are more frequently used to directly model *time* dependencies. The same applies to *feedbackward* architectures, since they require special base networks in order to work properly. We rarely see *convolution* or *interlayer* architectures in time series analysis, although recent publications suggest that they are used more frequently now, due to their ability to reduce dimensionality. *Hybrid methods* are used more often with classic

ANN based systems rather than with other base networks, since the added method most of the time replaces the complex and different *no ANN* network.

4.2 Dependencies of neural networks and time series characteristics

Since in most cases from our literature review *hybrid methods* or special networks (*no ANN*) are used to model time series features we depict the correlation between hybrid modeling and no ANN with time series characteristics in Fig. 3. We clearly see that with most of the features like chaos or non-stationary present, hybrid models are far more often used, since standard ANNs are in some cases not able to achieve satisfactory results by themselves but rather need another additional method that either precedes or succeeds the application of the neural network. A special focus should also be given to the time architectural features of ANNs (e.g., LSTM, RNN) and their over proportional use when it comes to non-stationarity, chaos and especially noise. As in the classical models like GARCH or ARIMA, the noise is most likely autocorrelated and therefore has to be dealt with accordingly or otherwise it will be included into the model and will introduce a prediction bias. Considering the two advanced ANN classes “*no ANN*“ and “*hybrid method*“, it is apparent from the descriptive analysis, that some special treatment of time series data is needed and that the data cannot simply be processed by an out-of-the-box algorithm. When we look into the detailed algorithm classes of either class, we find a variety of different methods like Ridge Polynomial or Abductive Networks as well as fuzzy logic operators used as hybrid functions with classical ANNs that do not correlate with time series features in this small literature sample.

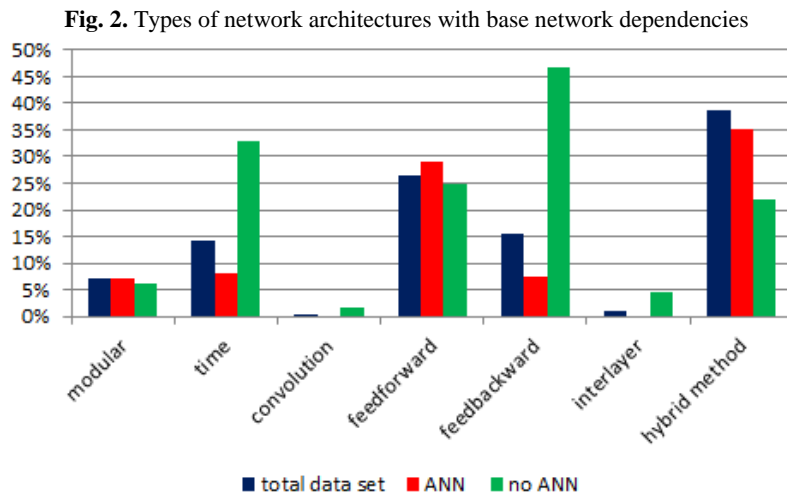
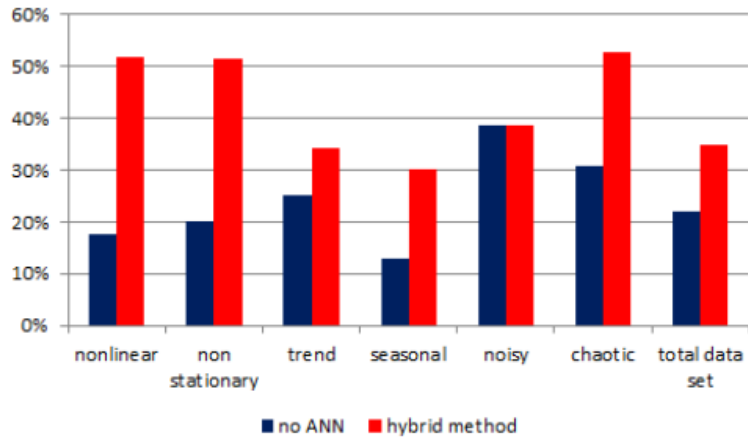


Fig. 3. Time Series Characteristics and ANN model

5 Summary and outlook

Since time series data plays an increasingly large role in economic applications not only in the finance sector but also in the context of event and process as well as failure data in the context of Industry 4.0, the knowledge base is in need of guidelines on how to design networks in order to handle certain types of time series with different characteristics. Towards that notion, we conducted an extended literature review with regard to time series characteristics and neural network architecture. We showed that several of those characteristics are of great concern in the literature and that certain hybrid or non-base ANNs are built for special cases. However, there is no correlation between certain types of characteristics and those algorithms. This can be explained by the special designs of hybrid functions and networks, such as wavelet networks that are specifically designed and configured to deal with one particular problem. In general the transparency on how different time series characteristics can be dealt with in the context of neuronal networks remain foggy. As this work continues it will conduct extensive benchmarking studies based on simulated and real time events in order to find useful correlations between time series characteristics and neural network architecture in the context of prediction problems.

6 References

1. Liu, W., Zidong;Liu, Xiaohui;Zeng, Nianyin;Liu, Yurong;Alsaadi, Fuad E.: A survey of deep neural network architectures and their applications. *Neurocomputing*. 234, 11-26 (2017).
2. Tu, J.V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* 49, 1225-1231 (1996).
3. Schmidhuber, J.: Deep learning in neural networks: An overview. *Neural Netw.* 61, 85–117 (2015).
4. Långkvist, M., Lars;Loutfi, Amy: A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit. Lett.* 42, 11-24 (2014).
5. Tkáč, M., Robert: Artificial neural networks in business: Two decades of research. *Appl. Soft Comput.* 38, 788-804 (2016).
6. Brockwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting*. Springer International Publishing (2016).
7. Franses, P.H., Van Dijk, D.: Forecasting stock market volatility using (nonlinear) GARCH models. *J. Forecast.* 229–235 (1996).
8. Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. *Phys. Rev. Lett.* 59, 845 (1987).
9. Mann, M.E., Lees, J.M.: Robust estimation of background noise and signal detection in climatic time series. *Clim. Change.* 33, 409–445 (1996).
10. Fettke, P.: State-of-the-Art des State-of-the-Art - Eine Untersuchung der Forschungsmethode Review innerhalb der Wirtschaftsinformatik. *Wirtschaftsinformatik.* 48, 257–266 (2006).
11. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a MIS Q. 26, 13–23 (2002).
12. Nickerson, R.C. ;Varshne., Upkar;Muntermann, Jan: A method for taxonomy development and its application in information systems. *Eur. J. Inf. Syst.* 22, 336-359 (2013).