Contents

1 Introduction

2 Literature Review

Feedforward Neural Networks Artificial neural networks (ANN) date back to the 40s where McCulloch and Pitts (1943) published their logical calculus for nervous activity stating that for "any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes". Especially Feedforward Neural Networks (FNN) are one of the most commonly used yet simplest ANN architectures. They usually exhibit a layered structure of activation functions where the information flows from the input layer to the subsequent hidden and output layers. Particularly multilayer perceptrons (MLP) with nonlinear actiavtion functions are used for a wide rage of applications.

Bartlmae and Rauscher (2000) successfully apply a FNN for risk management and forecasting daily volatilities. They are able to demonstrate a that FNNs underestimate risk less often than GARCH models. Anders, Korn, and Schmitt (1998) as well as Meissner and Kawano (2001) find that especially the MLP delivers a significantly better option pricing performance that the well known model by Black and Scholes (1973) when using i.a. GARCH volatility forecasts as inputs. L. Yu, Lai, and Wang (2006) use general regression neural networks (GRNN) to forecast shocks in the south east asian economonies. By framing the experiment as a classification problem, they find that GRNNs improve forecasts of regressions techniques such as logit and probit models. A recent study by Arnerić, Poklepović, and Teai (2018) finds that a hybrid approach using a FNN heterogeneous autoregressive (HAR) model to forecast realized variance yield better in sample results but does not improve out of sample accuracy. Using hybrid ANN architectures for volatility forecasting has proven to be useful in other studies as well (Kristjanpoller and Minutolo, 2018; Sallehuddin and Hj. Shamsuddin, 2009).

One major shortfall of FNNs, however, is the fact that they map input data directly to the output labels when learning. That makes it challenging to learn patterns in the context of time series since FNNs do not exhibit any memory state and consider every training example in an isolated way.

Recurrent Neural Networks The first versions of Recurrent Neural Networkswere developed by Hopfield (1982). They exhibit an architecture that feeds the output of every node to the other nodes of the same layer. Later Elman (1990) and Jordan (1986) developed that idea further and introduced returning connections to the node itself as well as connections to nodes in other layers in the network. Through this artificial lag, this architecture is well suited for learning sequence data and predicting time series.

Tino, Schittenkopf, and Dorffner (2001) use an Elman RNN with one hidden layer for volatility trading. They notice that their RNN behaves like a limited memory source and hardly beats classical fixed order Markov models. Dunis and Huang (2002), however, use an slightly diffrent architecture that loops back from the output instead of the hidden layer. They find that a FX option straddle trading strategy based on their RNN forecast for historical volatility is able to beat the GARCH benchmark in the out-of-sample period. Bekiros and Georgoutsos (2008) use a volatility based RNN to predict direction of change in the market. They find that this approach can especially improve trading performance during bear markets. Vejendla and Enke (2013) find that RNNs exhibit a lower mean squared error when predicting historical volatility than FNNs do. Other approaches use hybrid models or sentiment analysis of news data in order to predict the volatility of stocks (Liu et al., 2017).

When it comes to time series, it is very challenging for RNNs to to learn long term dependecies when being trained with gradient descent (Bengio, Simard, and Frasconi, 1994). The reason for that is that during backpropagation the multiplication of the Jacobian matrices leads to vanishing or exploding gradients which results in unreasonable weight updates (Pascanu, Mikolov, and Bengio, 2013).

Echo State Networks Echo state networks (ESN, Jaeger, 2001) and liquid state machines (LSM, Maass, Natschläger, and Markram, 2002)are a sub categegory of reservoir computing. They can be seen as extensions of traditional neural networks that are capable of learning dynamical systems and temporal patterns. An ESN has many similarities to an RNN where the weights of the hidden layers are not subject to training though. This creates a reservoir that acts as a complex nonlinear dynamic filter which is capable of transforming the input signals with a high dimesional temporal mapping (Schrauwen, Verstraeten, and Van Campenhout, 2007). As a consequence, the training process is simplified to a linear regression problem yet the ability to learn complex mapping functions remains.

Even though ESNs have some very feasible properties for time series forecasting, their application in finance is still in the early stage. Lin, Yang, and Song (2009), however, find that ESNs outperform convetional ANN architectures when it comes to short term stock price predictions. Their benchmark includes a FNN, an Elman RNN and a ANN with radial basis functions as activation function (see Broomhead and Lowe, 1988). Grigoryeva et al. (2014) assess the performance of ESN when forecasting conditional realized variances. They find that especially paralell configurations with multiple parallel reservoirs exhibit good forecasting results.

Long Short Term Memory Neural Networks Long Short Term Memory Neural Networks (LSTM) belong to the familiy of RNNs as well where the activation functions are replaced by gated recurrent units (GRU) . By using a system of gate functions and a cell state, Hochreiter and Schmidhuber (1997) could sucessfully create models that exhibit long term memory abilities. That way, LSTMs can effectivly adress the vanishing gradient problem and learn dependencies in time series that lie further apart in time. S. Yu and Li (2018) find that a LSTM exhibit lower error measures than GARCH models when forecasting volatiliy. Especially long term dependencies are better captures by the LSTM model. Kim and Won (2018) develop hybrid LSTM networks that use multiple (E)GARCH models as input for forecasting volatility of the KOSPI 200 Index. They find that especially hybrid architectures can improve forecasts siginficantly.

3 Formalised Model Architectures

3.1 Feedforward Neural Networks

3.2 Recurrent Neural Networks

3.3 Echo State Networks

3.4 Long Short Term Memory Neural Networks

4 Methodology

4.1 Data

4.2 Model Set Up and Benchmarks

4.3 Implementation

5 Empirical Results

6 Conclusion

Nomenclature

ANN	Artificial Neural Network
ESN	Echo State Network
FNN	Feed Forward Neural Network
GARCH	Generalized AutoRegressive Conditional Heteroscedasticity
GRNN	General Regression Neural Network
GRU	Gated Recurrent Unit
HAR	Heterogeneous AutoRegressive
LSM	Liquid State Maschine
LSTM	Long Short Term Memory Neural Network
MLP	Multilayer Perceptron
RNN	Recurrent Neural Network

References

Anders, Ulrich, Olaf Korn, and Christian Schmitt (1998). "Improving the pricing of options: A neural network approach". In: *Journal of forecasting* 17.5-6, pp. 369–388.

Arnerić, Josip, Tea Poklepović, and Juin Wen Teai (2018). "Neural Network Approach in Forecasting Realized Variance Using High-Frequency Data". In: *Business Systems Research Journal* 9.2, pp. 18–34.

Bartlmae, Kai and Folke A Rauscher (2000). "Measuring DAX market risk: a neural network volatility mixture approach". In: *Presentation at the FFM2000 Conference, London*. Vol. 31.

Bekiros, Stelios D and Dimitris A Georgoutsos (2008). "Direction-of-change forecasting using a volatility-based recurrent neural network". In: *Journal of Forecasting* 27.5, pp. 407–417.

Bengio, Yoshua, Patrice Simard, and Paolo Frasconi (1994). "Learning long-term dependencies with gradient descent is difficult". In: *IEEE transactions on neural networks* 5.2, pp. 157–166.

Black, Fischer and Myron Scholes (1973). "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3, pp. 637–654.

Broomhead, David S and David Lowe (1988). *Radial basis functions, multi-variable functional interpolation and adaptive networks*. Tech. rep. Royal Signals and Radar Establishment Malvern (United Kingdom).

Dunis, Christian L and Xuehuan Huang (2002). "Forecasting and trading currency volatility: An application of recurrent neural regression and model combination". In: *Journal of Forecasting* 21.5, pp. 317–354.

Elman, Jeffrey L (1990). "Finding structure in time". In: *Cognitive science* 14.2, pp. 179–211.

Grigoryeva, Lyudmila et al. (2014). "Stochastic nonlinear time series forecasting using time-delay reservoir computers: performance and universality". In: *Neural Networks* 55, pp. 59–71.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hopfield, John J (1982). "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the national academy of sciences* 79.1, pp. 2554–2558.

Jaeger, Herbert (2001). "The echo state approach to analysing and training recurrent neural networks-with an erratum note". In: *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* 148.34, p. 13.

Jordan, Michael (1986). "Attractor dynamics and parallelism in a connectionist sequential machine". In: *Proc. of the Eighth Annual Conference of the Cognitive Science Society (Erlbaum, Hillsdale, NJ), 1986*.

Kim, Ha Young and Chang Hyun Won (2018). "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models". In: *Expert Systems with Applications* 103, pp. 25–37.

Kristjanpoller, Werner and Marcel C Minutolo (2018). “A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis”. In: *Expert Systems with Applications* 109, pp. 1–11.

Lin, Xiaowei, Zehong Yang, and Yixu Song (2009). “Short-term stock price prediction based on echo state networks”. In: *Expert systems with applications* 36.3, pp. 7313–7317.

Liu, Yifan et al. (2017). “Stock volatility prediction using recurrent neural networks with sentiment analysis”. In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, pp. 192–201.

Maass, Wolfgang, Thomas Natschläger, and Henry Markram (2002). “Real-time computing without stable states: A new framework for neural computation based on perturbations”. In: *Neural computation* 14.11, pp. 2531–2560.

McCulloch, Warren S and Walter Pitts (1943). “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133.

Meissner, Gunter and Noriko Kawano (2001). “Capturing the volatility smile of options on high-tech stocks?A combined GARCH-neural network approach”. In: *Journal of economics and finance* 25.3, pp. 276–292.

Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio (2013). “On the difficulty of training recurrent neural networks”. In: *International Conference on Machine Learning*, pp. 1310–1318.

Sallehuddin, Roselina and Siti Mariyam Hj. Shamsuddin (2009). “Hybrid grey relational artificial neural network and auto regressive integrated moving average model for forecasting time-series data”. In: *Applied Artificial Intelligence* 23.5, pp. 443–486.

Schrauwen, Benjamin, David Verstraeten, and Jan Van Campenhout (2007). “An overview of reservoir computing: theory, applications and implementations”. In: *Proceedings of the 15th European Symposium on Artificial Neural Networks. p. 471-482 2007*, pp. 471–482.

Tino, Peter, Christian Schittenkopf, and Georg Dorffner (2001). “Financial volatility trading using recurrent neural networks”. In: *IEEE Transactions on Neural Networks* 12.42, pp. 865–874.

Vejendla, Ajitha and David Enke (2013). “Evaluation of GARCH, RNN and FNN Models for Forecasting Volatility in the Financial Markets.” In: *IUP Journal of Financial Risk management* 10.1.

Yu, Lean, Kin Keung Lai, and Shou-Yang Wang (2006). “Currency crisis forecasting with general regression neural networks”. In: *International Journal of Information Technology & Decision Making* 5.03, pp. 437–454.

Yu, ShuiLing and Zhe Li (2018). “Forecasting Stock Price Index Volatility with LSTM Deep Neural Network”. In: *Recent Developments in Data Science and Business Analytics*. Springer, pp. 265–272.