Investigation of Prediction Capabilities using RNN Ensembles

Nijole Maknickiene[1] and Algirdas Maknickas[2]

[1]Department of Financial Engineering, Vilnius Gediminas Technical University, Sauletekio al. 11, Vilnius, Lithuania

[2]Department of Information Technologies, Vilnius Gediminas Technical University, Sauletekio al. 11, Vilnius, Lithuania

Keywords: Prediction, EVOLINO, Financial Markets, Recurrent Neural Networks Ensembles.

Abstract: Modern portfolio theory of investment-based financial market forecasting use probability distributions. This investigation used a neural network architecture, which allows to obtain distribution for predictions. Comparison of the two different models - points based prediction and distributions based prediction - opens new investment opportunities. Dependence of forecasting accuracy on the number of EVOLINO recurrent neural networks (RNN) ensemble was obtained for five forecasting points ahead. This study allows to optimize the computational time and resources required for sufficiently accurate prediction.

1 INTRODUCTION

Neural networks and their systems are successfully used in forecasting. There are several factors that determine the predictive accuracy of the prediction - input selection, neural network architecture and the quantity of training data.

The paper (Kaastra and Boyd, 1996)is to provide a practical introductory guide in the design of a neural network for forecasting economic time series data.

An eight-step procedure to design a neural network forecasting model is explained including a discussion of trade offs in parameter selection, prediction dependence on number of iterations. In paper (Walczak, 2001), the effects of different sizes of training sample sets on forecasting currency exchange rates are examined. It is shown that those neural networks-given an appropriate amount of historical knowledge - can forecast future currency exchange rates with 60 percent accuracy, while those neural networks trained on a larger training set have a worse forecasting performance. More over, the higher-quality forecasts, the reduced training set sizes reduced development cost and time. In the paper (Zhou et al., 2002), the relationship between the ensemble and its component neural networks is analysed, which reveals that it may be a better choice to ensemble many instead of all the available neural networks. This theory may be useful in designing powerful ensemble approaches. In order to show the feasibility of the theory, an ensemble of twelve NN approach named GASEN is presented.

The methodology in paper (Tsakonas and Dounias, 2005) proposes an architecture-altering technique, which enables the production of highly antagonistic solutions while preserving any weight-related information. The implementation involves genetic programming using a grammar-guided training approach, in order to provide arbitrarily large and connected neural logic network. The ensemble of 1-5 neural networks was researched by (Nguyen and Chan, 2004), resumed that "incorporating more neural networks into the model does not guarantee that the error would be lowered. As it can be seen in the application case study, the model with two neural networks did not perform more satisfactorily than the single neural network.

In paper (Garcia-Pedrajas et al., 2005), was proposed a general framework for designing neural network ensembles by means of cooperative coevolution. The proposed model has two main objectives: first, the improvement of the combination of the trained individual networks; second, the cooperative evolution of such networks, encouraging collaboration among them, instead of a separate training of each network. Authors (Siwek et al., 2009)made ensemble of neural predictors is composed of three individual neural networks. The experimental results have shown that the performance of individual predictors was improved significantly by the integration of their results. The improvement is observed even during the application of different quality. In paper (Uchigaki et al., 2012) a prediction technique was proposed which was called "an ensemble of simple regression models" to improve the prediction accuracy of cross-project pre-

Maknickienė N. and Maknickas A..
Investigation of Prediction Capabilities using RNN Ensembles.
DOI: 10.5220/0004554703910395
In *Proceedings of the 5th International Joint Conference on Computational Intelligence* (NCTA-2013), pages 391-395
ISBN: 978-989-8565-77-8

diction. To evaluate the performance of the proposed method, was conducted 132 combinations of cross-project prediction were conducted using datasets of 12 projects from NASA IVV Facility Metrics Data Program. (Brezak et al., 2012)made a comparison of feed-forward and recurrent neural networks in time series forecasting. The obtained results indicate satisfactory forecasting characteristics of both networks. However, recurrent NN was more accurate in practically all tests using less number of hidden layer neurons than the feed-forward NN. This study once again confirmed a great effectiveness and potential of dynamic neural networks in modelling and predicting highly nonliner processes.

In order to form the investment strategies in financial markets, there is a need for a proper forecasting technique, which can forecast the future profitabilities of assets (stock prices or currency exchange rates) not as particular values but as probability distributions of values. Such approach is analytically meaningful because future is always uncertain and we cannot make any unambiguous conclusion about it. For this reason the adequate portfolio model is used, developed by (Rutkauskas, 2000), which is an amplification of Markowitz portfolio model. The adequate portfolio conception is based on the adequate perception of reality that portfolio return possibilities should be expressed as a probability distribution with its parameters. The analysis of the whole probability distribution is especially important taking into account that portfolio return possibilities usually do not conform to Normal probability distribution form and therefore it is not enough to know their mean value and standard deviation. The initial concept of adequate portfolio over time was also applied to the analysis of other complex processes in the scientific works of A.V.Rutkauskas and his coauthors (Rutkauskas et al., 2008), (Rutkauskas and Lapinskait-Vvohlfahrt, 2010), (Rutkauskas and Stasytyte, 2011), (Rutkauskas, 2012).

Decision maker is soliciting opinions as data for statistic inference, with the additional complication of strategic manipulation from interested experts (Riley, 2012). Authors investigated proportion of correct decisions made by different number of agents - 1; 1000.

The aim of the paper is to investigate the influence of the number of neural nets on accuracy of financial markets prediction, to find new conditions of constructing investment portfolios. Knowing how much RNN is enough that the ensemble makes sufficiently accurate forecasting, to allow the saving of time and power resources.

2 PREDICTION USING ARTIFICIAL INTELLIGENCE

The forecasting we understand the ability to correctly guess a certain amount of unknown data in time with some precision. After all, the predicted data set is compared with a set of known data to evaluate the correlation between these. Suppose it is known that p is an element of some set of distributions P. Choose a fixed weight w_q for each q in P such that the w_q add up to 1 (for simplicity, suppose P is countable). Then construct the Bayesmix $M(x) = \sum q w_q q(x)$, and predict using M instead of the optimal but unknown p. How wrong could this be? The recent work of Hutter provides general and sharp loss bounds (Hutter, 2007): Let $LM(n)$ and $Lp(n)$ be the total expected unit losses of the M-predictor and the p-predictor, respectively, for the first n events. Then $LM(n)Lp(n)$ is at most of the order of $\sqrt{Lp(n)}$. That is, M is not much worse than p. And in general, no other predictor can do better than that. In particular, if p is deterministic, then the M-predictor won't make any more errors. If P contains all recursively computable distributions, then M becomes the celebrated enumerable universal prior. The aim of this paper is to construct a model that can make predictions with a small enough difference $M(t)p(t)$ for some fixed time t. Autors (Schmidhuber et al., 2005), (Wierstra et al., 2005)propouse a new class of learning algorithms for supervised recurrent neural networks - RNN EVOLINO. EVolution of recurrent systems with Optimal LINear Output. EVOLINO- based LSTM recurrent networks learn to solve several previously unlearnable tasks. "EVOLINO-based LSTM was able to learn up to 5 sines, certain context-sensitive grammars, and the Mackey-Glass time series, which is not a very good RNN benchmark though, since even feed-forward nets can learn it well" (Schmidhuber et al., 2007). Modularity is a feature often found in nature. It can be of two types-1) when the modules are connected to each other in parallel or sequentially, 2) when the modules are connected by another module inside. We constructed a modular EVOLONO RNN system connecting them in parallel. It is very important to investigate an accuracy of prediction when new models are testing. The comparison of the performance of the forecasting models was made in terms of the accuracy of the forecasts on the test case domain.

3 DESCRIPTION OF MODELS BASED ON ENSEMBLES OF RECURRENT NEURAL NETWORKS

Two different models have been developed and tested. Technical feasibility has been a major factor in determining both the creation of models.

3.1 Points based Prediction Model

Evolino RNN-based prediction model, which is applied to the average for PC. This model, which uses eight predictors, was investigated with the phython program by the following steps:

Data step. Getting historical financial markets data from Meta Trader - Alpari. We choose for prediction EUR/USD (Euro and American Dollar), GBP/USD (Great Britain Pound and American Dollar), exchange rates and their historical data for the first input, and for the second input, two years historical data for XAUUSD (gold prise in USA dollars), XAGUSD (Silver price of USA dollars), QM (Oil price in USA dollars), and QG (Gass price in USD dollars). At the end of this step we have a basis of historical data.

Input step. The python script calculates the ranges of orthogonality of the last 80140 points of the exchange rate historical data chosen for prediction, and an adequate interval from the two years historical data of XAUUSD, XAGUSD, QM, and QG. A value closer to zero indicates higher orthogonality of the input base pairs. Eight pairs of data intervals with the best orthogonality were used for the inputs to the Evolino recurrent neural network. Influence of data orthogonality to accuracy and stability of financial market predictions was described in paper (Maknickas and Maknickiene, 2012).

Prediction Step. Eight Evolino recurrent neural networks made predictions for a selected point in the future. At the end of this step, we have eight different predictions for one point of time in the future.

Consensus Step. The resulting eight predictions are arranged in ascending order, and then the median, quartiles, and compatibility are calculated. If the compatibility is within the range [0; 0.024], the prediction is right. If not, then step 3 is repeated, sometimes with another "teacher" if the orthogonality is similar. At the end of this step, we have one most probable prediction for the chosen exchange rate.

Decision of trading are making by constructing portfolio of exchange rates with taking into account of predictions - medians, got by described model.

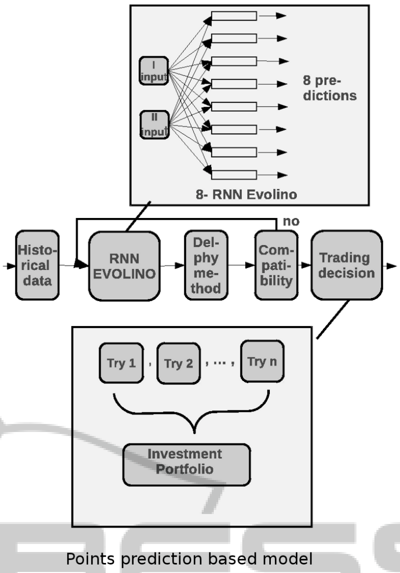

Figure 1: Scheme of point prediction based model.

3.2 Distribution based Predictions Model

Second - Evolino RNN-based prediction model. For calculation of big amount of ensembles software and hardware acceleration were employed. Every predicting neural network from ensemble could be calculated separately So, calculations could be done in parallel. MPI wrapper mpi4py (Dalcin, 2012)were used for this purpose. Cycle of each predicting neural network was divided into equal intervals and every interval were calculated on separate processor node. There are not needs for communication between mpi threads, so obtained equal to one efficiency of parallelism, where efficiency is described as folow (Fox et al., 1988), (Kumar et al., 1994):

$$S = \frac{1}{P}\frac{T_{seq}}{T(P)}, \quad (1)$$

where P is number of processors, $T(P)$ is the runtime of the parallel algorithm, and T_seq is the runtime of the sequential algorithm. Hardware acceleration were achieved using six nodes of Intel(R) Xeon(R) CPU E5645 @ 2.40 GHz on the cloud www.time4vps.eu. So calculations of ensemble of 300 predicting neural networks are 6.25 hours time long.

The first two steps - Data step and Input step - remain the same as in the first model. This is followed by other steps:

Prediction Step. We can choose n neural network forecasting. Neural networks can lead to the number of hours required for a decision. Therefore, it is

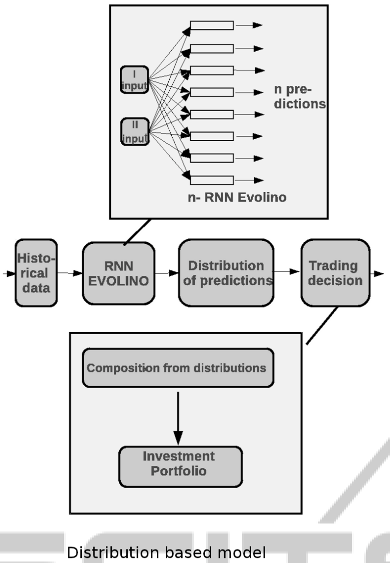

Figure 2: Scheme of distribution based model.

necessary to select the optimum number of ensemble. When $n > 60$, the forecast assumes the shape of the distribution. At the end of this step, we have a distribution with all parameters of it - mean, median, mode, skewness, kurtosis and et. Decision of trading are making by composed portfolio of exchange rates by analysing the distribution parameters.

4 COMPARISON OF PREDICTIONS ACCURACY

The test of the accuracy of models on 1-5 steps ahead 6 forecasts was investigated by MAPE. An interval forecast is considered to be correct if the actual value falls in side the predicted 95 % confidence interval. Point estimation accuracy was measured using the Mean Absolute Percentage Error (MAPE) of forecasts:

$$P_{ea} = 100 - \frac{100}{N} \sum_{i} \frac{|Y_i - \hat{Y}_i|}{Y_i}, \qquad (2)$$

where N - number of observations in the test set, Y_i - actual output and $\hat{Y}_i$ - forecasted output. Test from 5 observations was made in 20/01/2012 - 15/03/2012 (Fig. 3) Accuracy of predictions obtained in the interval 94-99,6%. Increase of accuracy depends on number of networks and forecasting becomes more stable. This investigation shows that in some cases more is not always better - with a lot of predictions EVOLINO RNN require more calculating processes time and resources. An interval of number of EVOLINO RNN [1; 100] has hight accuracy, but is not stable. Distribution of predictions has not form of clear shape and parameters are not informative. An interval [100; 200] is accurate and stable, so it not require too many time and resources. Distribution of predictions is sufficiently informative. An interval of n [200; 300] is good for investigation, but require to many calculation time - the investment decision in finance market so could be too late.

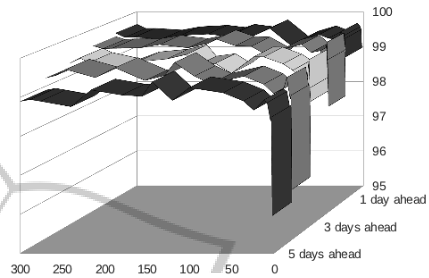

Figure 3: Dependency forecasting accuracy of the number of RNN EVOLINO: a) in 1 and 2 days ahead; b) 3, 4 and 5 days ahead.

1 and 2 points ahead forecasts are accurate and stable, and 3, 4 and 5 points ahead forecasts stability is reached only when the ensemble consists of over 64 RNN. In time series forecasting, the magnitude of the forecasting error increases over time, since the uncertainty increases with the horizon of the forecast.

5 CONCLUSIONS

Neural network architecture is very important in the forecasting process. The single neural network system provides a point forecast that accuracy is very unstable. Ensemble from eight neural networks provides more accurate forecasting point in the expected range. When number of neural networks exceeds 120, obtained distribution of predictions, which opens up new opportunities for investment portfolio opportunity. However, more is not always better. The ensemble for prediction requires more calculating time and resources. Stable and not feather growing prediction accuracy, gotten by increasing the number of RNN in ensemble, when $n > 120$, allows to optimize the investment decision-making process.

Those ensembles makes it possible to expect prediction accuracy of up to 5 days into the future. The decision to invest in the financial markets are always taken under uncertainty. Therefore, distributions are more informative and more reliable than the scatter projections. Application of distributions of probabilities in the investment portfolio needs further investigation.

REFERENCES

Brezak, D., Bacek, T., Majetic, D., Kasac, J., and Novakovic, B. (March, 2012). A comparison of feed-forward and recurrent neural networks in time series forecasting. In *Computational Intelligence for Financial Engineering Economics (CIFEr), 2012 IEEE Conference on*, pages 1–6.

Dalcin, L. (12/01/2012). https://code.google.com/p/mpi4py/.

Fox, G., Johnson, M., Lyzenga, G., Otto, S., Salmon, J., and Walker, D. (1988). Solving problems on concurrent processors vol 1. *1373849*.

Garcia-Pedrajas, N., Hervas-Martinez, C., and Ortiz-Boyer, D. (June, 2005). Cooperative coevolution of artificial neural network ensembles for pattern classification. *Evolutionary Computation, IEEE Transactions on*, 9(3):271–302.

Hutter, M. (2007). On universal prediction and bayesian confirmation. *Theoretical Computer Science*, 384(1):33–48.

Kaastra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3):215–236.

Kumar, V., Gmma, A., and Anshul, G. (1994). Introduction to parallel computing: Design and analysis of algorithms.

Maknickas, A. and Maknickiene, N. (2012). Influence of data orthogonality to accuracy and stability of financial market predictions. In *The Fourth International Conference on Neural Computation Theory and Applications (NCTA 2012)*, pages 616–619, Barcelona, Spain.

Nguyen, H. and Chan, C. (2004). Multiple neural networks for a long term time series forecast. *Neural Computing & Applications*, 13(1):90–98.

Riley, B. (2012). Practical statistical inference for the opinions a.of biased experts blake riley. In *In: The 50th Annual Meeting of the MVEA*, volume 2. Missouri Valley Economic Association.

Rutkauskas, A. and Lapinskait-Vvohlfahrt, I. (2010). Marketing finance strategy based on effective risk management. In *The Sixth International Scientific Conference Business and Managemen*, volume 1, pages 162–169. Technika.

Rutkauskas, A., Miečinskien, A., and Stasytyt, V. (2008). Investment decisions modelling along sustainable development concept on financial markets. *Technological and Economic Development of Economy*, 14(3):417–427.

Rutkauskas, A. V. (2000). Formation of adequate investment portfolio for stochasticity of profit possibilities. *Property management*, 4(2):100–115.

Rutkauskas, A. V. (2012). Using sustainability engineering to gain universal sustainability efficiency. *Sustainability*, 4(6):1135–1153.

Rutkauskas, A. V. and Stasytyte, V. (2011). Optimal portfolio search using efficient surface and three-dimensional utility function. *Technological and Economic Development of Economy*, 17(2):305–326.

Schmidhuber, J., Wierstra, D., Gagliolo, M., and Gomez, F. (2007). Training recurrent networks by evolino. *Neural Computation*, 19(3):757–779.

Schmidhuber, J., Wierstra, D., and Gomez, F. (2005). Modeling systems with internal state using evolino. In *In Proc. of the 2005 conference on genetic and evolutionary computation (GECCO)*, pages 1795–1802, Washington. ACM Press, New York, NY, USA.

Siwek, K., Osowski, S., and Szupiluk, R. (2009). Ensemble neural network approach for accurate load forecasting in a power system. *International Journal of Applied Mathematics and Computer Science*, 19(2):303–315.

Tsakonas, A. and Dounias, G. (2005). An architecture-altering and training methodology for neural logic networks: Application in the banking sector. In Madani, K., editor, *Proceedings of the 1st International Workshop on Artificial Neural Networks and Intelligent Information Processing, ANNIIP 2005*, pages 82–93, Barcelona, Spain. NSTICC Press. In conjunction with ICINCO 2005.

Uchigaki, S., Uchida, S., Toda, K., and Monden, A. (Aug.2012). An ensemble approach of simple regression models to cross-project fault prediction. In *Software Engineering, Artificial Intelligence, Networking and Parallel Distributed Computing (SNPD), 2012 13th ACIS International Conference on*, pages 476–481.

Walczak, S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. *Journal of management information systems*, 17(4):203–222.

Wierstra, D., Gomez, F. J., and Schmidhuber, J. (2005). Evolino: Hybrid neuroevolution / optimal linear search for sequence learning. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI),*, pages 853–858, Edinburgh.

Zhou, Z.-H., Wu, J., and Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1):239–263.