Review of Nature-inspired Forecast Combination Techniques

Christiane Lemke and Bogdan Gabrys
Computational Intelligence Research Group, School of Design, Engineering & Computing
Bournemouth University, Poole House, Talbot Campus
Poole, BH12 5BB United Kingdom
Phone: +44-1202-965491, Fax: +44-1202-595314
email:{clemke, bgabrys}@bournemouth.ac.uk

ABSTRACT: Effective and efficient planning in various areas can be significantly supported by forecasting a variable like an economy growth rate or product demand numbers for a future point in time. More than one forecast for the same variable is often available, leading to the question whether one should choose one of the single models or combine several of them to obtain a forecast with improved accuracy. In the almost 40 years of research in the area of forecast combination, an impressive amount of work has been done. This paper reviews forecast combination techniques that are nonlinear and have in some way been inspired by nature.

KEYWORDS: Forecast Combination, Nonlinear Forecasting, Adaptive Forecasting

INTRODUCTION

Since the publication of the seminal paper on forecast combination in 1969 ([1]), research in this area has been active; recent reviews and summaries can be found in [2] and [3]. In general, four main reasons for the potential benefits of forecast combinations have been identified:

1. It is implausible to be able to correctly model the true data generation process using only one model. Single models are most likely to be simplifications of a much more complex reality, so different models might be complementary to each other and be able to approximate the true process better.
2. Even if a single best model is available, a lot of specialist knowledge is required in most cases to find the right functions and parameters. Forecast combinations help achieve good results without in-depth knowledge about the application and without time consuming, computationally complex fine-tuning of a single model.
3. It is not always feasible to take all the information an individual forecast is based on into account and create a superior model, because information may be private, unobserved or provided by a closed source.
4. Individual models may have different speeds to adapt to changes in the data generation process. Those changes are difficult to detect in real time, which is why a combination of forecasts with different abilities to adapt might perform well.

Most of the work that has been done focuses on the linear combination of forecasts, in which the combined forecast $\hat{y}^c$ is calculated as the weighted sum of m individual forecasts $\{\hat{y}_1,..,\hat{y}_m\}$ as shown in the following equation:

$$\hat{y}^c = \sum_{i=1}^{m} \omega_i \hat{y}_i \qquad (1)$$

Weights can be estimated in various ways, one easy and often remarkably robust example is the simple average combination with equal weights. More complex algorithms calculate weights using regression or based on the past performance of the individual models.

In [3], it is suspected that the reason for the small number of publications on nonlinear forecast combination lies in the parameter estimation errors, which are already large for linear combination schemes and can be expected to be even larger for nonlinear ones that come with a higher degree of freedom in the estimation process. However, nonlinear models are more likely to capture the true data generation process by combining the individual forecasts with a nonlinear function ψ :

$$\hat{y}^c = \psi\left[\hat{y}_1, \hat{y}_2, .., \hat{y}_m\right] \quad (2)$$

In the next chapters, nonlinear forecast combination methods that use nature-inspired algorithms including neural networks, fuzzy systems, genetic programming, self-organising algorithms and neuro-fuzzy systems are summarised and discussed.

NEURAL NETWORKS

Artificial neural networks have been inspired by neural processing in the human brain and are widely appreciated as data-driven universal approximators for arbitrarily complex functions. Consisting of a number of simple processors (neurons) joined by weighted connections that are trained by learning algorithms, complex relationships and patterns in the input data can be discovered.

The first use of neural networks for combining forecasts appears in [4]. The authors use a four layer feedforward network trained with the backpropagation algorithm on a time series data set of IBM stock prices with 120 values. Three individual forecasts obtained using one exponential smoothing, one trend analysis and one ARIMA model are combined. Diagrams imply that the neural network has been trained well and shows a good out-of-sample performance. Numerical results of the same experiment can be found in [5], where the neural network improves the mean squared error of the best individual model by 27%. The mean squared error of a linear combination method estimating weights according to past variance of individual models could be improved by 29%.

Less favourable results can be found in more extensive studies. In [6], a data set consisting of four macroeconomic series over almost 18 years is used. Two individual forecasts are combined in a neural network with one hidden layer that consists of four nodes using one linear and three nonlinear logistic transfer functions. Furthermore, two linear combinations are used for comparison, with weights estimated by ordinary least squares regression and past mean absolute deviation. Looking at a squared error measure, the neural network combinations perform at least as well as individual forecasts and similar to the least squares regression forecast combination. This is however not the case for an absolute error measure. Additionally, the authors point out that neural networks are the only model in the evaluation that was not encompassed by at least one other forecast, meaning that none of the other models can provide the same information and indicating superior forecast quality. The same experimental setup with similar results is used in [7], with the difference that some parameters of the neural network are determined using evolutionary programming.

Another empirical study can be found in [8]. Four individual forecasting and three forecast combination methods including the simple average, an ordinary least squares regression and a one layer neural network with four nodes have been evaluated on one and a half years of financial data including a crisis period. The neural network performs particularly well in the crisis period, but, in contrast to the results described in the previous paragraph, the simple average of individual forecasts or individual models themselves always outperform this approach if squared prediction errors are used for comparison. If the performance of the neural network is assessed by absolute error values, it gets significantly better, but it is still not always superior.

In [9], the authors explain potential benefits of the neural network approach graphically, plotting surfaces of the combining functions in relation to two input forecasts. While linear approaches create a flat surface with a constant response to changing input forecasts, a neural network shows a highly flexible response function where the impact of a change in one of the individual forecasts is influenced by the value the other forecast assumes at the same time. Neural network combination models can thus account for interactions between individual forecasts which would be omitted when using linear combination schemes.

FUZZY SYSTEMS

Uncertainty, vagueness and judgement are common elements of human reasoning. Fuzzy logic is a concept that attempts to formalise those elements by extending ordinary Boolean logic, allowing truth values between “completely true” and “completely false”.

Fuzzy systems for forecast combination can be found following two different paradigms. First, fuzzy systems can be seen as a kind of regime model, where two or more different forecasting models can be active at one time. In contrast, ordinary regime models switch abruptly from one set of model coefficients to another, dependent on the state the system is likely to be in. The authors of [10] follow this approach and use a first order Takagi-Sugeno fuzzy system, which basically means that the conclusions in the rules are represented by linear functions of the input variables. Parameters in the system are estimated using the gradient descent algorithm to minimise its squared forecast error. Tests are carried out using three time series with 200 to 312 values, combining one ARIMA and one K-nearest neighbours model. Looking at four different error measures, the fuzzy system almost always outperforms or draws level with the individual forecasts and linear forecast combination methods based on simple average or regression. The authors argue furthermore, that because of using a linear mixture of models, their resulting system is simpler than one working with neural networks. They admit at the same time, that the computational complexity of building their system is very high and suggest simplifications.

Two more publications emphasise a different aspect of fuzzy systems - the possibility of modelling linguistic and subjective knowledge. A case study for a fuzzy logic combination system for forecasting the demand for signal transmission products is presented in [11]. Membership functions are automatically generated by processing incoming data, and thus adapted if new data arrives. Combining a customer and a market expert's forecast for the demand rate with traditional simple forecasting methods like averaging previous values and trend, two periods of time are compared. The fuzzy forecast improves the root mean squared error of the customer forecast by 63.5% on average; results comparing the combined forecast to the other input forecasts are not given.

A similar hierarchical rule base combining subjective forecasts from experts with time series forecasts is presented in [12], the individual time series forecasts being provided by an ARMA and a decomposition forecasting model. Only a hypothetical example is used for evaluation, improving the individual time series forecast error by 22% and 46%, respectively. Noteworthy about the presented system is its periodically adapting rule base, which adjusts confidence in rules according to their past performance. Experimental results are given with and without this adaptive component, showing that the learning mechanism reduces the mean squared error from 7.1% to 5.8% per month.

EVOLUTIONARY COMPUTATION FOR POOLING

Genetic programming is a subfield of evolutionary computation which was inspired by Darwinian biological mechanisms in evolution such as mutation, recombination, natural selection, reproduction and survival of the fittest.

In [13], the authors look at forecast combination for airline ticket demand data. In this area, forecasts are available on different levels because of the complex network of flights, routings and itineraries which contain different fareclasses or points of sale. A large number of predictions can thus be generated on these various levels with different forecasting models. To avoid the drawbacks that arise from using too many input forecasts without taking the risk of choosing among them, the authors propose hierarchical structures for forecast combination on the multiple levels as shown in figure 1. Genetic programming is used to evolve the tree-like combination structures where individual forecasts are represented by leafs, different combination methods by nodes and the final combined forecast is obtained at the root. As the fitness function, i.e. the criterion to optimize, the mean absolute deviation forecasting error on a previously unseen test data set has been chosen. The initial population is either a random subset or selected by experts; the standard operators for genetic programming are applied in the following way: crossover exchanges two subtrees and mutation exchanges a node or leaf. Variations of this algorithm are assessed on demand rates for airline tickets, the best of them leads to consistent improvements in the mean absolute error, for example improvements of 6% compared to the simple average combination.

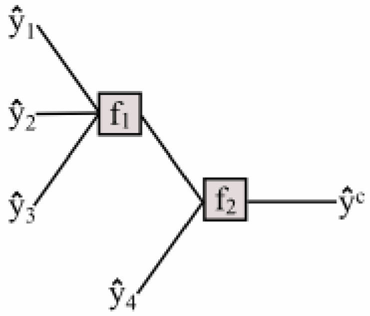

Figure 1: Example of a multilevel combination of individual forecasts $\{\hat{y}_1,..,\hat{y}_4\}$ with two functional approaches f_1, f_2, producing the combined forecast $\hat{y}^c$.

This approach is extended in [14], where the authors work with a large pool of individual forecasts for airline ticket demand that was generated with three diversifying procedures: the use of different function spaces, thick modelling and parameters learnt at different data aggregation levels. A large number of input forecasts can however lead to difficulties in the combination process, such as increased model complexity and increased parameter estimation errors. Pooling is a popular procedure to address these issues by grouping similar forecasts and subsequently combining the pooled forecast. Research in this area started only recently in [15], where forecasts are clustered based on recent past error variance. In [14], the authors follow theoretical evidence that combining forecasts differing in more than one diversification dimension can lead to non-optimal combination weights and decrease the combined forecast's accuracy. Consequently, they propose a hierarchical pooling structure, where initial pools are created with forecasts that only differ regarding one of the diversification approaches used. The forecasts in each cluster are trimmed to exclude the ones with the worst past performance and then linearly combined. In that way, resulting output forecasts are reduced by one diversification dimension and can again be pooled according to a different one. Combination structures are created using evolutionary computation, with the order of the dimensions being the optimisation problem. Different degrees of dynamics are investigated, i.e. allowing created pools to be split once again according to past error variance of individual forecasts or assigning different dimensions to be used for subsequent pooling to each intermediate pool. Experimental results are given in relation to the performance to the current commercial forecasting system used for airline data and indicate a significant improvement. Furthermore, a "flat" combination structure of five individual forecasts is outperformed by 9.8%.

SELF-ORGANISING ALGORITHMS

Self-organisation refers to the ability of a system to increase its complexity as a result of many relatively simple interactions, without being led by outside influences. It is a concept that can be observed in physics, chemistry and biology as well as in social sciences. Research on a popular inductive self-organising method called the Group Method of Data Handling (GMDH) was started in [16]. This algorithm aims at finding the structure of a model by generating candidates in an iterative process, sorting out possible solutions according to an external criterion in each step.

An application of this concept to forecast combination can be found in [17]. A statistically learning network is built, starting with individual forecasts as input variables. A transfer function, usually a polynomial, is used to generate a first layer consisting of a set of model candidates. Parameter estimation for the models takes place on training data, for example using regression. The resulting candidates are ranked according to an external criterion, which could be the mean squared forecast error on unseen test data. Only models that are able to improve the value of the external criterion are selected as inputs for the subsequent layer, where the generation, estimation and selection process starts again. The value of the external criterion will pass through a minimum, which is when the optimal model has been found. A sketch of a simple possible outcome of the described algorithm can be found in figure 2.

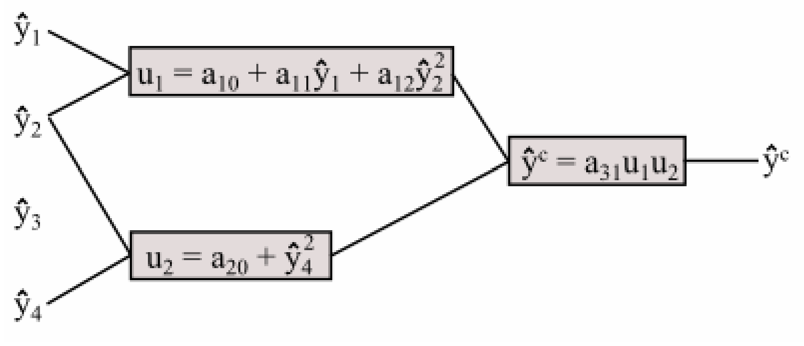

Figure 2: Example of a polynomial network generated by a self-organising algorithm, $\{\hat{y}_1,..,\hat{y}_4\}$ being individual input forecasts, u_1, u_2 intermediate models with the learnt parameters a_{ij}, producing the combined forecast $\hat{y}^c$.

Compared to neural networks, this approach has several advantages. It does not require time-consuming training and gives an explicit model of the system. Furthermore, the optimal number of layers and neurons does not need to be determined in advance, which is a problem that still remains unsolved in the area of neural networks.

Only a very small empirical experiment is presented in [17] to evaluate the performance of the self-organising approach, just one macroeconomic series over 20 months is used. Compared to a neural network combining approach, all individual models and a linear least squares combination, the proposed algorithm outperforms every single one of them. The linear least squares method performed worst of all three combination methods in terms of out-of-sample performance measured by MSE. In comparison to this method, artificial neural networks decreased the mean squared error by 13%; and the network generated by self-organising algorithm reduced it even further by 26%.

COMPARATIVE STUDIES

One comparative study investigating neural networks, fuzzy systems and neuro-fuzzy systems for forecast combination has been found [18]. It uses a data set consisting of 226 observations of temperature readings of a chemical plant; for which two forecasts are generated in-sample with an ARMA model and Holt-Winter's smoothing algorithm. For the evaluation of the nonlinear combination methods, the data set is divided into a training set of 150 and a testing set of 76 values. As the first model, a feedforward neural network with two hidden layers containing six neurons each is trained to combine the two forecasts using the Levenberg-Marquardt algorithm. As the second model, a fuzzy system with symmetric Gaussian membership functions and automatically generated rules is employed. Finally, a neuro-fuzzy system is created, which combines the learning ability of the neural network approach with the interpretability of the fuzzy system by learning parameters of the fuzzy system (mean and variance of the Gaussian functions and the centers of the fuzzy regions) with a neural network which is again trained by a Levenberg-Marquardt algorithm. Comparing algorithms using four different error measures, it is concluded that each of the presented nonlinear methods always outperforms both of the individual forecasts. The neuro-fuzzy approach performs best among the three, for example reducing the mean squared error of the best individual forecasting by 79%. It should however be noted that the individual forecasts have been evaluated only in-sample, which might not correctly reflect their behaviour in out-of-sample applications.

CONCLUSION

This paper identified and reviewed publications that use nonlinear nature-inspired methods for forecast combination, covering neural networks, fuzzy systems, self-organising algorithms, evolutionary computation and a neuro-fuzzy approach.

The most investigated of those methods is clearly forecast combination with neural networks. While some publications report a significant improvement over individual forecasting and linear forecast combination models, others find that performance is mixed dependent on forecasting horizons or the error measure used. Indications also exist that neural networks perform better than other methods if there are structural breaks in the data generation process. The mixed results and the traditional drawbacks of neural networks might favour self-organising approaches, although only one publication mentioned them in the context of forecast combination, so they do require further investigation. Fuzzy systems are the method to use if subjective or linguistic forecasts have to be included in the combination. If this is not the case, the computationally complex process of generating and applying a fuzzy system can provide a reason against it. However, some publications on fuzzy systems also consider adaptivity to a changing environment thoroughly, which has not been done for any of the other methods yet. Evolved multilevel combination structures for a forecast pool generated by different diversifications have a solid theoretical base and work well in empirical experiments, but have only been applied to airline data so far.

Compared to the vast amount of literature on linear forecast combination, the number of publications about nonlinear, nature-inspired methods appears small. Empirical results are mostly smaller case studies; the majority only use one time series with less than 300 observations for assessment of algorithms. The choice of individual forecasts used for combination is different in every study, making it difficult to compare results. Evaluations across methods have only been found in one publication. The general opinion about nonlinear forecasting methods is that they lead to more unstable results than linear methods, because of more freedom in the parameter estimation. However, some of the reviewed papers do provide results that encourage further research in that area.

REFERENCES

[1] Bates, J. M.; Granger, C. W. J., 1996, "The combination of forecasts", Operational Research Quart., vol. 20, pp. 451 - 468.

[2] Granger, C. W. J., 2001, "Invited review: combining forecasts – twenty years later", Journal of Forecasting, vol. 8, pp. 167 – 173.

[3] Timmerman, A., 2005, "Forecast combinations", Handbook of Economic Forecasting, Elliot, G.; Granger, C. W. J.; Timmerman, A. (editors), pp. 135 – 196.

[4] Shi, S.; Liu, B., 1993, "Nonlinear combination of forecasts with neural networks", Proceedings of 1993 Joint Conference on Neural Networks, Nagoya, Japan, vol. 1, pp. 959 - 962.

[5] Shi, S.; Da Xu, L.; Liu, B., 1999, "Improving the accuracy of nonlinear combined forecasting using neural networks", Expert Systems with Applications, vol. 16, pp. 49 – 54.

[6] Donaldson, R.; Kamstra, M., 1996, "Forecast combining with neural networks", Journal of Forecasting, vol. 15, pp. 49-61.

[7] Harrald, P.; Kamstra, M., 1997, "Evolving artificial neural networks to combine financial forecasts", IEEE Transactions on Evolutionary Computation, vol. 1, pp. 40 – 52.

[8] Hu, M.; Tsoukalas, C., 1999, "Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates", Journal of International Financial Markets, Institutions and Money, vol. 9, pp. 407 - 422.

[9] Donaldson, R.; Kamstra, M., 1999, "Neural network forecast combining with interaction effects", Journal of the Franklin Institute, vol. 336, pp. 227 – 236.

[10] Fiordaliso, A., 1998, "A nonlinear forecast combination method based on Takagi-Sugeno fuzzy systems", International Journal of Forecasting, vol. 14, pp. 367 – 379.

[11] Frantti, T.; Mähönen, P., 2001, "Fuzzy logic-based forecasting model", Engineering Applications of Artificial Intelligence, vol. 14, pp. 189 – 201.

[12] Petrovic, D.; Xie, Y.; Burnham, K., 2006, "Fuzzy decision support system for demand forecasting with a learning mechanism", Fuzzy Sets and Systems, vol. 157, pp. 1713 – 1725.

[13] Riedel, S.; Gabrys, B., 2005, "Evolving multilevel forecast combination models – an experimental study", Proceedings of NiSIS 2005 Symposium, Albufeira, Portugal.

[14] Riedel, S.; Gabrys, B., 2007, "Dynamic pooling for the combination of forecasts generated using multi level learning", Proceedings of the International Joint Conference on Neural Networks, Orlando, Florida

[15] Aiolfi, M.; Timmerman, A., 2006, "Persistence in forecasting performance and conditional combination strategies", Journal of Econometrics, vol. 127, pp. 31 – 53.

[16] Ivakhnenko, A.G., 1970, "Heuristic self-organization in problems of engineering cybernetics", Automatica, vol. 6, pp. 207 - 219.

[17] He, X.; Xu, X., 2005, "Combination of forecasts using self-organizing algorithms", Journal of forecasting, vol.24, pp. 269 – 278.

[18] Palit, A.; Popovic, D., 2000, "Nonlinear combination of forecasts using artificial neural network, fuzzy logic and neuro-fuzzy approaches", Proceedings of the ninth IEEE International Conference on Fuzzy Systems, San Antonio, Texas, USA, pp. 566 - 571.