



USING A MIXTURE OF EXPERTS' APPROACH TO SOLVE THE FORECASTING TASK

Victor SINEGLAZOV, Elena CHUMACHENKO, Vladyslav GORBATYUK

*Technical University of Ukraine "KPI", 03056 Peremogy Ave 37, Kiev, Ukraine
E-mail: lobach21@mail.ru (corresponding author)*

Received 12 November 2013; accepted 20 August 2014



Victor SINEGLAZOV, Prof. Dr Eng.

Date and place of birth: 1950, Odessa, Ukraine.

Education: the National Technical University of Ukraine "KPI".

Affiliation and functions: Doctor of Technical Sciences (Eng), Professor, Senior Member IEEE, has been working at the National Aviation University since 1989, Director of the Institute of Electronics and Control Systems at the National Aviation University, Head of the Department of Aviation Computer-Integrated Systems since 1998, active member of the International Academy of Navigation and Control and of the New York Academy of Sciences.

Research interests: automated control systems and information processing.

Publications: over 485 publications.



Elena CHUMACHENKO, Assoc. Prof.

Date and place of birth: 1950, Kutaisi, Georgia.

Education: Georgian Polytechnic Institute.

Affiliation and functions: PhD, Associate Professor, has been working at the Department of Technical Cybernetics at the National Technical University of Ukraine "KPI" since 1976.

Research interests: modeling and management of flexible computer systems.

Publications: over 35 publications.



Vladyslav GORBATYUK

Date and place of birth: 1992, Ilyichyivsk, Ukraine.

Education: the National Technical University of Ukraine "KPI".

Affiliation and functions: 5th year student at the National Technical University of Ukraine "KPI" since 2009.

Research interests: artificial neural networks, forecasting methods.

Publications: 5 articles.

Abstract. The forecasting problem appears frequently in the aviation industry (demand forecasting, air transport movement forecasting, etc.). In this article, a new approach based on multiple neural networks of different topologies is introduced. An algorithm was tested on real data and showed better results compared to several other methods. This shows its suitability for further usage in aviation forecasting tasks.

Keywords: artificial neural network, forecasting method, mixture of experts, mean square prediction error, group method of data handling, training set.

1. Introduction

The forecast problem arises in almost every area of human activity, including the aviation industry. One important task in the aviation industry is demand forecasting, which is one of the most crucial issues of inventory management. The high cost of modern aircraft and the expense of such repairable spares as aircraft engines and avionics constitute a large part of the total investment of many airline operators. An insufficient amount of produced aircraft can lead to excessive delay costs while their overproduction will result in downtime.

Forecasting techniques in aviation have grown more sophisticated over the years and are widely used in aviation nowadays. Owing to the stochastic nature of demand for aircraft, airline operators perceive difficulties in making predictions and are still looking for good forecasting methods.

Many methods have been developed to solve the task (and new methods are constantly evolving), starting from a simple linear regression (Radchenko 2011) and ending with complex neural networks (Gioqinang, Hu 1998) and hybrid systems (Bodyanskiy et al. 2008). Due to the variety of different methods, it is often a difficult question which model to select. One approach was suggested by the author of a group method of data handling (GMDH) (Madala, Ivakhnenko 1994), which relied on defining a set of considered models, finding the parameters for each model and selecting the best models according to some external criteria. This approach has greatly evolved since then, and it has been shown to be very effective for solving real-life problems. In this paper, another possible approach to the forecasting problem which uses an external criterion to weight models (as well as GMDH) is offered.

It is suggested that this method may further be applicable for use in the demand forecasting task for regional aviation facilities and other industrial sectors which have similar demand patterns to those of airlines.

2. The mathematical formulation of the problem of forecasting

Let us have discrete n samples $\{x_1, x_2, \dots, x_n\}$ at successive time points t_1, t_2, \dots, t_n . Then the problem of prediction (Fig. 1) consists of the prediction of the value x_{n+k} at some future point of time t_{n+k} where k is the duration of the forecast:

$$x_{n+k} = F(x_1, x_2, \dots, x_n), \tag{1}$$

where F is some unknown function.

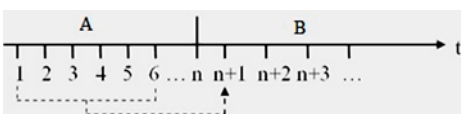


Fig. 1. Graphical illustration of the problem statement forecasting: A indicates known values, B – the forecast period

3. Review of existing methods of forecasting

The method of moving average (Radchenko 2011). The method is based on a simple model which assumes that the current value of a number

$$y_t, (t=1 \dots n), \tag{2}$$

of the series

$$\{y_1, y_2, \dots, y_n\}, \tag{3}$$

is the average sum of a number of previous values y_1, y_2 and some random component.

The weighted moving average method (Alesinskaya 2002). The next step is the modification of the model using an assumption that more recent values reflect the situation more accurately. This leads to assigning weights to previous values, hence more recent values will be assigned greater weights.

Group method of data handling (Ivakhnenko 1968). GMDH is a set of forecasting algorithms which is based on splitting the original data into two sets: for training and testing, and the usage of some kind of base functions the parameters of which are derived from the training set, and a test of how well they simulate a given raw is performed on the test sample.

Artificial neural networks (ANN) (Amir, Samir 1999). ANN is distributed, adaptive, nonlinear learning machines based on an information processing model that emulates human brain activity. The basic unit (also called the processing element, PE) has input and output parameters similar to a biological neuron. An error obtained by comparing the output of the network with the desired output is used as a feedback to adjust the weights connecting the processing elements. The multilayer perceptron (MLP) is the most commonly used network. ANN are a good model for a forecasting task because of their nonlinearity; hence, they can model complex dependencies between input variables and a forecasted output variable, and have the possibility to learn.

The application of ANN (namely time-lagged feed-forward neural networks) for the prediction of passenger traffic flows was described by T. O. Blinova (2007) and P. Kozik, J. Şep (2012) and they showed reasonably good results.

Another algorithm based on ANN and GMDH for the aircraft overhaul demand forecasting task was mentioned in (Sineglazov et al. 2013) and it showed better results compared to ANN on their own.

4. Suggested approach

As it was mentioned previously, artificial neural networks have been successfully applied to the forecasting task. For instance, in (Mohsen, Yazdan 2007; Khan, Ondrusek 2002), a MLP topology was successfully used,

while in (Jerome *et al.* 1994) an Elman neural network (ENN) was involved. The sensible question is: what topology to use? It seems that currently there is no certain answer. In (Jerome *et al.* 1994), it was shown that the MLP topology is better for autoregressive-like processes (Box *et al.* 1994), while the ENN is more suited for autoregressive-moving average processes (Box *et al.* 1994). The main idea of the proposed approach is to consider multiple topologies at the same time. The algorithm consists of the following steps:

1. Normalize the source time series using the following normalization technique: $x_i^n = \frac{x_i - \bar{x}}{\sigma_x}$.
2. Preprocess the normalized time series using Tukey 53H algorithm (Klevecka, Lelis 2008). This algorithm was developed to remove outliers from the data, and it is very important, since outliers can cause a big change in a model's parameters. Of course, another smoothing technique, such as wavelet decomposition can be used (Akansu, Liu 1991).
3. Take a preprocessed time series and transform it into a training sample matrix using a sliding window technique. Therefore, if we have n points, and the prehistory size equals l , and the forecasting period is k ($l > 0, k > 0, l + k \leq n$), then we will have $n - l - k + 1$ training samples.
4. Split the obtained matrix of samples according to a certain ratio (usually 0.7:0.3) into the training and validating sets. How to do this "well" is also an interesting question, which won't be discussed here. For reference, it was discussed in (Ivakhnenko, Iurachkovskiy 1987). We suggest using a random division of samples.
5. Train three different neural networks using the training set. The networks' topologies are: MLP, ENN and radial-basis function (RBF) network (Broomhead, Lowe 1988). Again, a lot of articles were published regarding issues of a suitable architecture of MLP, ENN or RBF networks together with questions regarding suitable choice of their parameters, such as activation functions, and the learning algorithm. We suggest the following parameters:
 - The architecture of every network is l input neurons (since there are l inputs, there is no other choice), l neurons in the hidden layer and one output neuron. If, after going through all steps, the resulting model is still unsatisfactory, the number of neurons in the hidden layer should be multiplied by a factor of 2, and the whole procedure should be repeated.
 - The activation function of neurons in the hidden layer (applies only to the ENN and MLP, since RBF networks have predefined activation functions) is

a hyperbolic tangent $\tanh(x) = \frac{e^{a^*x} - e^{-a^*x}}{e^{a^*x} + e^{-a^*x}}$; the activation function of neurons in the output layer is linear.

- The learning algorithm for the MLP is Levenberg-Marquardt (Levenberg 1944), for the ENN - gradient descent with momentum and adaptive learning rate backpropagation.
6. Train another, so-called "gating" network, using the validating set (Jacobs *et al.* 1991).
 - The network parameters:
 - Topology – RBF network;
 - Architecture – l input neurons, l neurons in the hidden layer and three neurons in the output layer.
 - Activation functions of the output layer – sigmoid $f(x) = \frac{1}{1 + e^{-a^*x}}$.
 - The network is trained as follows (given sample \vec{s}_i in the validating set).
 - *Step 1.* Feed the sample to each of the already trained networks. Three forecasts should be obtained: $\hat{y}_i = [\hat{y}_{i1}, \hat{y}_{i2}, \hat{y}_{i3}]^T$, where \hat{y}_{i1} is the output of a MLP, \hat{y}_{i2} – ENN, \hat{y}_{i3} – RBF network.
 - Calculate the relative error of each forecast using an actual value y_i :

$$\delta_i = [\delta_{i1}, \delta_{i2}, \delta_{i3}]^T, \delta_{ij} = \left| \frac{\hat{y}_{ij} - y_i}{y_i} \right|, j = 1 \dots 3.$$
 - Then normalize the calculated vector, so that its elements sum to one.
 - Finally, a desired output of a gating network for the input sample \vec{s}_i , given the vector of relative errors, received on the previous step, is calculated as $out = [\frac{1 - \delta_{i1}}{2}, \frac{1 - \delta_{i2}}{2}, \frac{1 - \delta_{i3}}{2}]$.
 - Simply put, the gating network is responsible for predicting the weights for the forecasts of other three networks, given the input sample. I.e., for the samples in the validating set that are better predicted by MLP it should give a bigger weight to MLP, while giving smaller weights to other networks' forecasts and vice versa. That is why a desired output vector is constructed as one minus the relative error of each network. Therefore, the smaller the error the bigger the weight which should be assigned to the network's forecast for this sample.
 7. After all the networks are trained, the final output of the constructed model, given a vector of inputs \vec{x} is calculated as follows:

$$\hat{y} = gate(\vec{x})_1 * MP(\vec{x}) + gate(\vec{x})_2 * ENN(\vec{x}) + gate(\vec{x})_3 * RBF(\vec{x}),$$
 where the outputs of a gate network should be normalized, so that they sum to one (another

option is to use the softmax activation function in the output layer of a gate network) (Sutton, Barto 1998). If the obtained model is unsatisfactory (its forecasts are not accurate enough), the whole procedure should be repeated but with a doubled number of neurons in the hidden layer of each network.

5. Application of the proposed algorithm

For the testing of the proposed algorithm a public set of aircraft sales (the total number of aircraft sold in U.S. per year) from 1947 to 2011 was used. The data was split into 3 sets: for training, validating and testing as shown in figure 2.

The prediction results obtained after training the model using the suggested approach are shown in figure 3.

The MSE value on the testing set equals 0.0194, when comparing it with the value obtained by ANN – 0.0613, and via a model proposed in (Sineglazov et al. 2013) – 0.0255. We can conclude that this model is more suitable for this particular case.

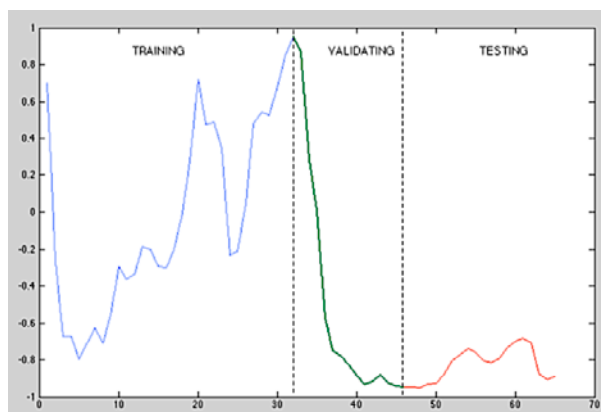


Fig. 2. Splitting the source data into 3 sets

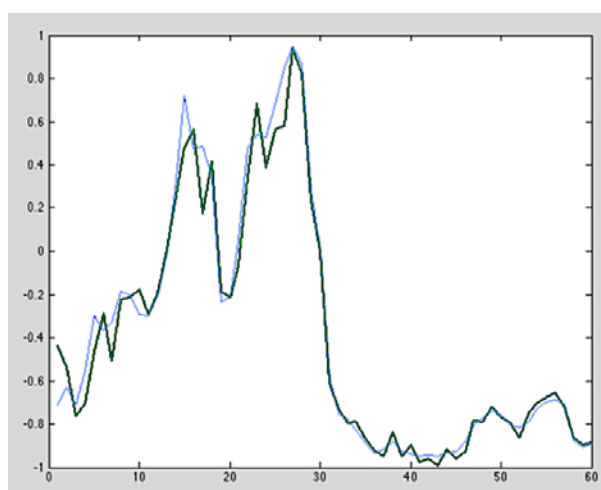


Fig. 3. The prediction results: blue indicates source data, green – forecast

6. Conclusions

The proposed method is another variation of a “mixture of experts” approach, where the considered experts are neural networks with different topologies. This is done to eliminate the big problem of topology selection of a neural network for a given data: instead of selecting a single topology, we train three different topologies and then train a so-called “gating” network, which will be responsible for weighting forecasts of the trained “forecasting” networks. To train a gating network a new validating set is used, i.e. we are actually using an external criterion, similar to GMDH algorithms.

The proposed algorithm has shown better results in comparison with several other forecasting methods, which means that it is suitable for further usage in forecasting tasks.

References

- Akansu, A. N.; Liu, Y. 1991. On signal decomposition techniques, *Optical Engineering Journal* 30(7): 912–920. <http://dx.doi.org/10.1117/12.55886>
- Alesinskaya, T. V. 2002. *Metodi skolziashchego srednego i eksponencialnogo sglagivaniya*: Uchebnoe posobie po resheniu zadach po kursu “Ekonomiko-matematicheskie metodi i modeli”. Taganrog: TRTU. 153 p. (in Russian).
- Amir, F. A.; Samir, I. S. 1999. A comparison between neural-network forecasting techniques – case study: river flow forecasting, *IEEE Transactions on Neural Networks* 10(2): 402–409. <http://dx.doi.org/10.1109/72.750569>
- Blinova, T. O. 2007. Analysis of possibility of using neural network to forecast passenger traffic flows in Russia, *Aviation* 11(1): 28–34.
- Bodyanskiy, Ye.; Pliss, I.; Vynokurova, O. 2008. Adaptive wavelet-neuro-fuzzy network in the forecasting and emulation tasks, *International Journal on Information Theory and Applications* 15(1): 47–55.
- Box, G.; Jenkins, G. M.; Reinsel, G. C. 1994. *Time series analysis: forecasting and control*. 3rd ed. Prentice-Hall.
- Broomhead, D. S.; Lowe, D. 1988. *Radial basis functions, multi-variable functional interpolation and adaptive networks*: Technical report. 4148. RSRE.
- Gioqinang, Z.; Hu, M. Y. 1998. Neural network forecasting of the British pound/US dollar exchange rate, *International Journal of Management Science* 26: 495–506.
- Ivakhnenko, A. G. 1968. Metod gruppovogo urakhivaniya argumentiv – konkurent metodu stokhastichnoi approssimacii, *Avtomatika* 3: 58–72. (in Russian).
- Ivakhnenko, A. G.; Iurachkovskiy, Iu. P. 1987. *Modelirovanie slozhnyh sistem po eksperimentalnim dannim*. Moscow: Radio i sviaz. (in Russian).
- Jacobs, R. A.; Jordan, M. I.; Nowlan, S. J., et al. 1991. Adaptive mixtures of local experts, *Neural Computation* 3: 79–87. <http://dx.doi.org/10.1162/neco.1991.3.1.79>
- Jerome, T. C.; Douglas, R. M.; Atlas, L. E. 1994. Recurrent neural networks and robust time series prediction, *IEEE Transactions on Neural Networks* 5(2): 240–254. <http://dx.doi.org/10.1109/72.279188>
- Khan, M. R.; Ondrusek, C. 2002. Short term load forecasting with multilayer perceptron and recurrent neural networks, *Journal of Electrical Engineering* 53(1–2): 17–23.

- Klevecka, I.; Lelis, J. 2008. Pre-processing of input data of neural networks: the case of forecasting telecommunication network traffic, *Telektronikk* (3/4): 168–178.
- Kozik, P.; Şep, J. 2012. Aircraft engine overhaul demand forecasting using ANN, *Management and Production Engineering Review* 3(2): 21–26.
- Levenberg, K. 1944. A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics* 2: 164–168.
- Madala, H. R.; Ivakhnenko, A. G. 1994. *Inductive learning algorithms for complex systems modeling*. London, Tokyo: CRC Press.
- Mohsen, H.; Yazdan, S. 2007. Artificial neural network approach for short term load forecasting for Illam region, *Engineering and Technology* 1(4): 280–284.
- Radchenko, S. G. 2011. *Metodologiya regressionnogo analiza: Monographiya*. K.: Korniyuchuk. 376 p. (in Russian).
- Sineglazov, V.; Chumachenko, E.; Gorbatyuk, V. 2013. An algorithm for solving the problem of forecasting, *Aviation* 17(1): 9–13. <http://dx.doi.org/10.3846/16487788.2013.777219>
- Sutton, R. S.; Barto, A. G. 1998. *Reinforcement learning: an introduction*. Cambridge: The MIT Press.