

Ensembles of Neural Network for Telemetry Multivariate Time Series Forecasting

Alexander Doudkin¹⁾, Yauheni Marushko¹⁾

1) United Institute of Informatics Problems of National Academy of Sciences of Belarus, Belarus, Minsk, st. Surganova 6, marushkoe@gmail.com, uiip.bas-net.by/structure/1_is/index.php

Abstract: In this paper, we propose an approach to solving the problem of forecasting multivariate time series of telemetry data using ensemble neural networks. Approaches to the construction of neural network ensembles are analyzed and accuracy of prediction is evaluated. It is studied a possibility of training the neural network ensembles for reducing errors of multivariate time series forecasting.

Keywords: Forecasting, artificial neural network, ensemble of neural networks, telemetry, multivariate time series.

1. INTRODUCTION

Space telemetry is a set of technologies that allows remote collection of information about on-board spacecraft subsystems. The subsystems are controlled by analysis of sensor readings that are distributed across submodules. A subsystem state at a particular point in time is described by a vector of sensor values. The time sequence of states is a sequence of vectors of sensor values. Hence, space telemetry data are multivariate time series.

One of the tasks of analysis is the forecasting of this time series.

The task of forecasting multivariate time series is generally stated as follows [1, 2]: from the known current value of the sequence $y(k)$ and the prehistory $y(k-1), y(k-2), \dots, y(k-m)$ we should evaluate the next value $\hat{y}(k+1)$. Each element of the sequence $y(k)$ represents a vector of values at time k having length m , which is used for prediction called time window.

A variety of techniques has been used in short-term forecasting, including regression analysis and “time series” analysis. Simple regression and multiple linear regression are frequently used forecasting methods. They have an advantage that they are comparatively simple and can easily be implemented. However, they are somewhat limited in their ability to forecast in certain situations, especially in the presence of nonlinear relationships and high levels of noisy data. Most “time series” models belong to the class of linear time series forecasting, because they postulate a linear dependency of the future value on the past values. The most popular univariate models are the autoregressive moving average ARMA model and its derivatives. But the artificial neural networks (NN) often outperform these models in solving complicated tasks [1]. Deep neural networks can also be used [3], but in this case, the large training set is needed.

The processing and analysis of the telemetry data in a continuous process is accompanied by non-deterministic noises. In this case it is preferred to use of technology of artificial neural networks. The effectiveness of this technology depends on the neural network architectures

and learning methods [1, 4], which requires multiple experiments.

There are examples of using neural networks in on-board intelligent decision support systems for managing complex dynamic objects and diagnosis of its condition [5, 6].

In this paper, we investigate the possibility of short-term prediction parameters of telemetry using a neural network ensemble (ENN) [7, 8], which is a set of neural networks, and makes decisions by averaging the results of the separate neural networks.

Predictive analytics and machine learning are often faced with the “concept drift”, which means that the statistical properties of the target variable, which the model tries to predict, change over time in an unpredictable manner [9], this increases the prediction error. Hence, the neural network prediction efficiency can be improved by using iterative learning methods [7, 9, 10]. These methods involve accuracy estimation of the models and their ranking on each analysis iteration. In the case of lowering the overall accuracy, the ensemble detects the concept drift and a new member, trained on the relevant data is added to the ensemble. In this approach, the model laid during initial training is retained and the new parameters are entered without “forgetting” problem. Thus, an additional learning the neural networks ensemble is realized.

2. METHODS AND ALGORITHMS

The main objective of this study is to explore the possibility of using ensembles of neural networks for telemetry data forecasting. Therefore, neural networks and ensembles selected as the test models. Classic feedforward network with one hidden layer selected as a single neural network model. Ensembles with a different way of calculating the output values are analyzed.

A comparative analysis of the following approaches to the formation of the output value of the ensemble is performed:

1) The output value is formed as the sum of the individual networks outputs. It is calculated for the case with a single output neuron by the formula:

$$y = \frac{1}{n} \sum_{i=1}^n y_i, \quad (3)$$

where n – the number of the networks in the ensemble, y_i – the output of i -th network;

2) The output value is formed as a weighted sum of the individual networks outputs. It is calculated for the case with a single output neuron by the formula:

$$y = \sum_{i=1}^n y_i \cdot w_i, \quad (4)$$

where n – the number of the networks in the ensemble, y_i – the output of i -th network, w_i – the weight of i -th network, which is formed according to the formula:

$$w_i = \frac{mse_i}{\sum_{i=1}^n mse_i}, \quad (5)$$

where mse_i – MSE of i -th network on a validation set;

3) The output value is formed as the weighted sum of the outputs of the individual networks (formula (4-5)) and weighing is repeated after a certain interval of time samples with the evaluation on this interval (dynamically weighted ensemble).

Also an iterative method of ensemble learning was studied. The following algorithm was used for this purpose:

- 1) Processing of the current input vector.
- 2) Evaluation of the accuracy of the ensemble. For this purpose, errors in the previous step and the current are compared.
- 3) If the error is not increased or increased in a predetermined range, go to the next input vector.
- 4) Otherwise, the training set is formed, which includes all the accumulated data from the last additional training.
- 5) The formation and training of a new neural network.
- 6) Formed network is added to the ensemble.
- 7) Recalculation of the weighting coefficients based on their errors on the latest data produced for all neural networks of the ensemble.

3. EXPERIMENT

The data set is a finite set of precedents, which is selected in some way from the set of all possible precedents called the general population.

Data sets parameters for our experiments are presented in Table 1.

Table 1. Telemetry data of the correction propulsion system of the spacecraft

Title	Sampling time, s.	The dimension of the time series	Number of samples
Dt_set_s01	1	24	57501
Dt_set_s05	0.5	24	12245
Dt_set_sl	0.1	24	6613

Each set of telemetry data is obtained by sensors of the correction propulsion system. These include temperature parameters and pressure levels of the xenon supply unit, electric parameters of flow control, electrical parameters of the engines anode and cathode. Values provided by the sensors depend on the mode which is set by the control commands. Sensor values are correlated with each other; it allows to expect a satisfactory assessment of the forecast.

Since the learning is carried out with the teacher, it is necessary to form the learning set of pairs "input vector, output vector." Formation of the pair of the learning sample is carried out by windowing method [2]: it takes a certain period of time series and a few observations stand out from it, and that will be the input vector. The value of the desired output in the training example will be the next in order. Then, the window moves in the one position in the direction of increasing time, and the process of forming the next couple of training sample is repeated.

Thus, if the dimension of the time series data is N and the window size – W , the neural network should receive the input sample with size $N \times W$. So, for the window $W = 20$ Dt_set_sl set is converted into an input set with size of 6589×480 , and target set – with size 28×6589 .

Resampling and scaling performed during the preparation of the datasets of telemetric sensor information.

Resampling is performed to convert the raw data, representing as a sequence of time stamps of important events to the form with fixed sampling time. Scaling is necessary to bring the data in the valid range $[-1, 1]$. Also, the outputs of the network are scaled.

The aim of the experiment is to determine an influence of the parameters of a single neural network and ensembles on forecast performance.

Multilayer perceptron with one hidden layer with *hyperbolic tangent* non-linear activation function is used as the base element of the ensemble. RPROP algorithm is used to train a single network [11]. Prediction window is chosen to be 20 samples.

To evaluate the quality of trained neural networks as well as to compare different ensembles, the following values are used:

- mean square error, MSE:

$$MSE = \frac{1}{m} \sum_{i=1}^m e_i^2; \quad (1)$$

- mean absolute error, MAE:

$$MAE = \frac{1}{m} \sum_{i=1}^m |e_i|. \quad (2)$$

where $e_i = y_i - t_i$, y_i and t_i – the obtained and the desired output signals of i -th neuron of the output layer, respectively, m – the size of the output layer of the neural network.

In this experiment, the input set is divided in a ratio of 9:1 for a general training set and the final test set. The general training set is divided into validation set (15 %), a test set (15 %) and training set (70 %) randomly, which were used for training, evaluation and search for the best architecture, respectively.

The resulting test set is used to calculate the final estimates obtained by neural networks.

3.1 Evaluation of the hidden layer size

Suboptimal size of the hidden layer of the neural network was evaluated according to the following algorithm.

- 1) Determination of the search interval.
- 2) Training of 10 networks with the current hidden layer size, which is selected from the search interval.
- 3) Formation of the weighted ensemble.
- 4) Evaluation of the ensemble accuracy.
- 5) Until the end of the search range, go to the next element of the interval (step 2).
- 6) Select the ensemble with the lowest MSE on search range, the size of the hidden layer of the ensemble element is the best one.

The dependence of the accuracy of the hidden layer size is shown in Fig. 1.

The sizes of the hidden layers for the data sets are shown in the Table 2.

It should be noted that this algorithm significantly increases the formation of the ensemble.

Table 2. Evaluation of the hidden layer size

Dataset	Hidden layer size
Dt_set_s1	28
Dt_set_s05	33
Dt_set_s01	70

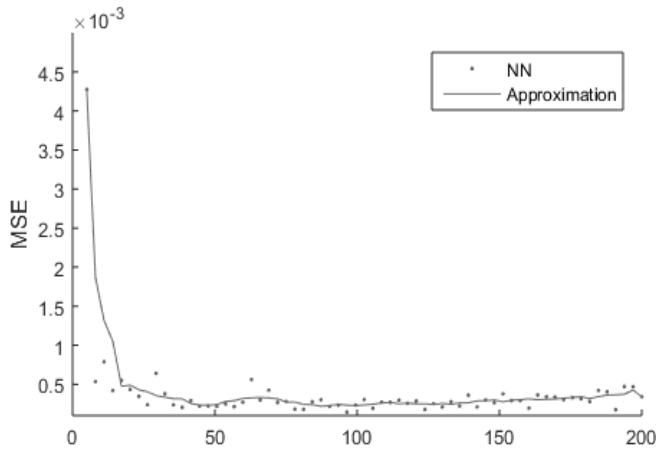


Fig. 1. – Evaluation of the hidden layer size for dataset Dt_set_s1.

3.2 Analysis of approaches for formation of the output value of an ensemble

MSE depends on the hidden layer size (it is shown in Fig. 2).

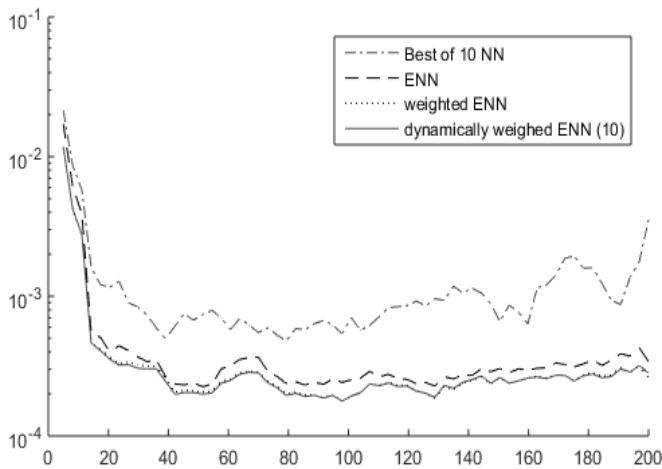


Fig. 2. – Different models error depending on the size of the hidden layer.

Evaluation of different models across Dt_set_s1 test set is given in Table. 3.

Table 3 shows that dynamically weighed ensemble has the smallest error and the difference in the estimated parameters for weighted ensembles is very small.

3.3 Evaluation of dynamically weighted of ensemble

Evaluation of weighting interval for dynamically weighed of the ensemble is performed. MSE plot for the various parameters is shown in Fig. 3.

Table 3. Evaluation of different models across Dt_set_s1 test set

Model	MSE,	MAE,
-------	------	------

	10^{-4}	10^{-3}
Single neural network	3.51	0.115
	3.66	0.114
Weighted ensemble of neural networks	2.76	9.65
Dynamically weighted ensemble of neural networks, interval = 10	2.75	9.59

The evaluation procedure includes the following steps:
1) Training the ensemble with the sub-optimal size of the hidden layer.

2) Determination of the search interval of weighting steps.

3) Evaluation of MSE for different models and the ensemble with the current step of weighting repetition.

4) Until the end of the search interval, go to the next element of the interval.

Fig. 3 shows that the dynamic weighting ensemble with a small step of repetitive weighting (less than 10 samples) has the smallest error.

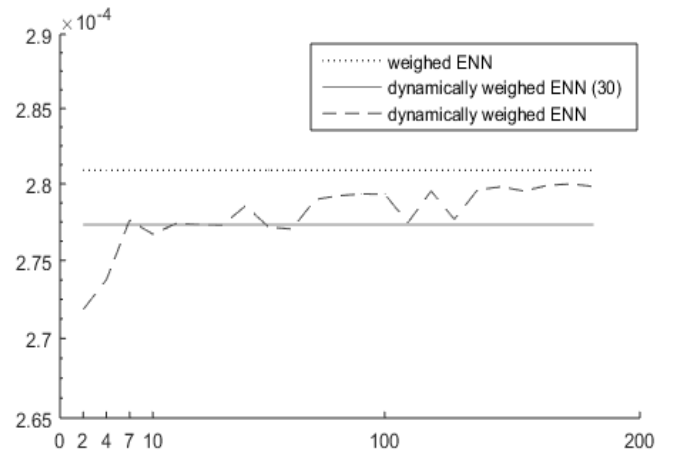


Fig. 3. – Evaluation of weighing interval for the dynamically weighted ensemble.

3.4 Evaluation in the case of concept drift

Concept drift refers to a change in value over time and, consequently, to a change in the distribution of the value. The environment from which these values are obtained is not stationary.

An expert ensemble which iteratively trained (without previous access to the data) on an input, in combination with some form of weighted voting for the final solutions is used in drift detection algorithms [9, 10].

For this experiment, modifications have been made artificially in the investigated data. The linearly increasing trend was added, and the sine wave signal was modeled as a periodic component.

Modification of Dt_set_s1 dataset is shown in Fig. 4. The first of the 1319 accounts are used without modification.

The evaluation procedure includes the following steps:

1) Train the ensemble.

2) Modify the data set by adding the trend and (or) the seasonal component.

3) Set the error threshold for additional training algorithm.

4) Set a minimum amount of data for accumulation.

5) Rate accuracy of different models, as well as an ensemble with additional training (additional training is performed only when the accumulated needed amount of data).

Evaluation of different architectures on a modified set of data shown in Table. 4.

Table 4. Evaluation of different architectures on a modified set

Model	MSE	MAE
Single neural network	0.6035	0.48428
Ensemble of neural networks	0.030388	0.12742
Weighted ensemble of neural networks	0.028874	0.12115
Dynamically weighted ensemble of neural networks, interval = 10	0.02661	0.11703
Dynamically weighted ensemble of neural networks, interval = 10 with additional training	0.015286	0.08767

All models showed a significant drop in the accuracy on a modified set, including the ensemble with additional training. This is associated with accumulation interval for the additional data set.

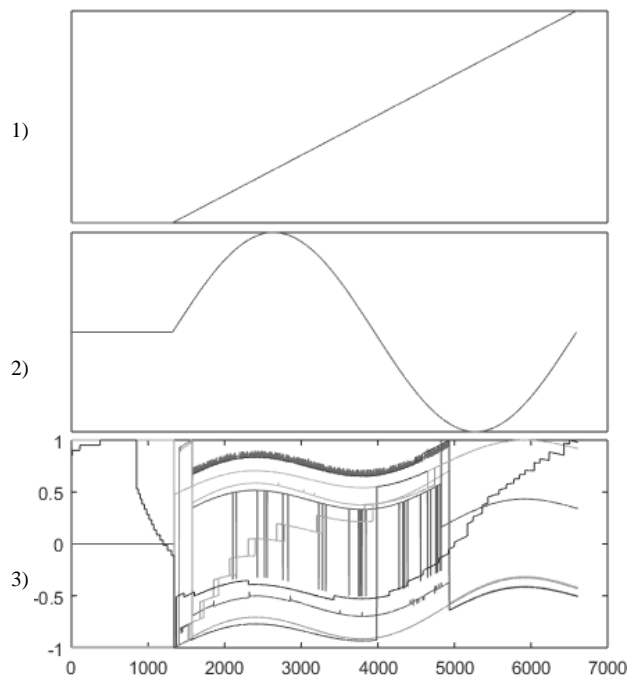


Fig. 4. – Modification of Dt_set_s1 data set: 1) Linearly increasing trend; 2) periodic component; 3) a modified multivariate signal.

4. CONCLUSION

Ensembles significantly reduce the forecasting error in comparison with single neural networks.

Dynamically weighted ensemble with a small step of repetitive weighing (less than 10 samples) has the smallest error and the difference in the estimated parameters for weighted ensembles is very small.

The minimum mean square error at short-term forecasting of telemetry data obtained by sensors of the correction propulsion system is equal to $2,75 \times 10^{-4}$.

All models showed a significant drop in the accuracy

on a modified set, including the ensemble with additional training. The ensemble of neural networks with additional training also showed a drop in precision due to needing of additional network training on enough samples of new data, but the resulting accuracy was higher.

5. REFERENCES

- [1] Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals / Hao Quan [et al.] // *IEEE Transactions on Neural Networks and Learning Systems*. – 2013. – Vol. 25, Iss. 2. – ISSN: 2162-237X. – P. 303-315.
- [2] Lysayak, A.S. Prediction of multidimensional time series / A.S. Lysayak, B.Y. Ryabko // *Vestnik SibGUTI*. – 2014. – № 4. – P.75–88. (in Russian)
- [3] Dalto, Mladen. Deep neural networks for time series prediction with applications in ultra-short-term wind forecasting / Mladen Dalto [et al.] // *IEEE International Industrial Technology*, Seville, Spain, March 17–19 th. – 2015.
- [4] Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir / M. Valipour [et al.] // *Journal of Hydrology* 476 (2013). – P. 433–441.
- [5] Khachumov V.M.. Review of Standards and the concept of monitoring, control and diagnostics of the spacecraft tools building / V.M. Khachumov [et al.] // – *Software Systems: Theory and Applications*. – №3(26), Volume 6. – 2015. – P.21-43. (in Russian)
- [6] Emelyanov Yu.G.: Neural orientation angles and distance of the spacecraft sensor control system / Yu.G. Emelyanov, K. A. Konstantinov, S. V. Pogodin, etc. // *Software Systems: Theory and Applications*, 2010. – № 1 (1). – P. 45-59. (in Russian)
- [7] Marushko, Y. Using Ensembles of Neural Networks with Different Scales of Input Data for the Analysis of Telemetry Data / Y. Marushko // *Proc. of the XV Intern. Ph.D. Workshop OWD 2013* (Wisla, 19–22 Oct. 2013). – Gliwice: Silesian University of Technology, 2013. – P. 386–391.
- [8] Neural network ensemble operators for time series forecasting / Nikolaos Kourentzes [et al.] // *Expert Systems with Applications*. – July 2014. – Vol.41, Iss. 9. – ISSN: 0957-4174. – P. 4235–4244.
- [9] Elwell, R. Incremental Learning of Variable Rate Concept Drift / R. Elwell, R. Polikar // *MCS*. – Vol. 5519 of Lecture Notes in Computer Science. – 2009. – P. 142–151.
- [10] Parikh, D. An ensemble-based incremental learning approach to data fusion / D. Parikh, R. Polikar // *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*. – 2007. – Vol. 37. – № 2. – P. 437–450.
- [11] Riedmiller, M. A direct adaptive method for faster backpropagation learning: The RPROP algorithm / M. Riedmiller, H. Braun // *In Proceedings of the IEEE International Conference on Neural Networks (ICNN)*, San Francisco. – 1993. – P. 586–591.