Metadata of the chapter that will be visualized in SpringerLink

Book Title	Pattern Recognition and Information Processing
Series Title	
Chapter Title	Ensembles of Neural Network for Telemetry Multivariate Time Series Forecasting
Copyright Year	2017
Copyright HolderName	Springer International Publishing AG

Author	Family Name	**Doudkin**
	Particle	
	Given Name	**Alexander**
	Prefix	
	Suffix	
	Division	
	Organization	United Institute of Informatics Problems of National Academy of Sciences of Belarus
	Address	st. Surganova 6, Minsk, Belarus
	Email	doudkin@newman.bas-net.by
Corresponding Author	Family Name	**Marushko**
	Particle	
	Given Name	**Yauheni**
	Prefix	
	Suffix	
	Division	
	Organization	United Institute of Informatics Problems of National Academy of Sciences of Belarus
	Address	st. Surganova 6, Minsk, Belarus
	Email	marushkoee@gmail.com

Abstract

In this paper, we propose to solve the problem of forecasting multivariate time series of telemetry data using neural network ensembles. Approaches to the forming neural network ensembles are analyzed and prediction accuracy is evaluated. The possibility of training the neural network ensembles is studied for reducing errors of multivariate time series forecasting.

Keywords (separated by '-')

Forecasting - Artificial neural network - Ensemble of neural networks - Telemetry - Multivariate time series

Ensembles of Neural Network for Telemetry Multivariate Time Series Forecasting

Alexander Doudkin and Yauheni Marushko(✉)

United Institute of Informatics Problems of National Academy of Sciences of Belarus,
st. Surganova 6, Minsk, Belarus
doudkin@newman.bas-net.by, marushkoee@gmail.com

AQ1

Abstract. In this paper, we propose to solve the problem of forecasting multivariate time series of telemetry data using neural network ensembles. Approaches to the forming neural network ensembles are analyzed and prediction accuracy is evaluated. The possibility of training the neural network ensembles is studied 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 analysis tasks is the forecasting of such time series.

The task of forecasting multivariate time series is generally formulated as follows [1, 2]: from the known current value of the sequence $y(k)$ and the prehistory $y(k{-}1), y(k{-}2), \ldots, y(k{-}m)$ 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. The length sequence m is called as time window.

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

V.V. Krasnoproshin and S.V. Ablameyko (Eds.): PRIP 2016, CCIS 673, pp. 1–10, 2017.
DOI: 10.1007/978-3-319-54220-1_6

The processing and analysis of the telemetry data is accompanied by non-deterministic noises. In this case it is preferred to use NNs technology. The effectiveness of this technology depends on NN architectures and learning methods [1, 4], which requires multiple experiments.

There are examples of using NNs 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 NNs and makes decisions by averaging the results of the separate NNs.

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] and 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 NN, 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 ENN is realized.

2 Methods and Algorithms

The main objective of this study is to explore the possibility of using ENNs for telemetry data forecasting. Therefore, NNs and ENNs are selected as the test models. Classic feed forward NN with one hidden layer is selected as a single NN model.

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:

$$y = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{1}$$

   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:

$$y = \sum_{i=1}^{n} y_i \cdot w_i, \tag{2}$$

   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:

$$w_i = \frac{mse_i}{\sum_{i=1}^{n} mse_i}, \tag{3}$$

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 (1–3)) 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. 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 expecting a satisfactory assessment of the forecast.

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_s1	0.1	24	6613

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_s1 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:

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m} e_i^2; \tag{4}$$

– mean absolute error, MAE:

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m} |e_i|. \tag{5}$$

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.

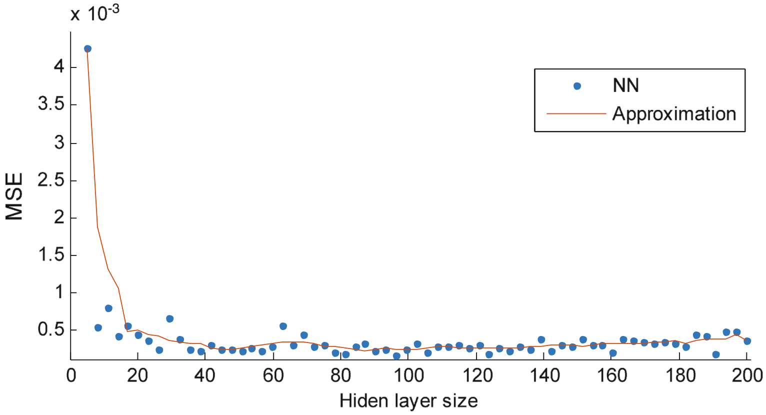

Fig. 1. Evaluation of the hidden layer size for dataset Dt_set_s1

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

Table 2. Evaluation of the hidden layer size

Dataset	Hidden layer size
Dt_set_s1	28
Dt_set_s05	33
Dt_set_s01	70

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

3.2 Evaluation of Neural Network Training Time

The aim of the experiment is to determine the differences in the speed of execution of training the neural network on a single central processing unit (CPU) and graphics processing unit (GPU).

Hardware for the experiment:
CPU Intel Core i5 4200H, 2 cores, 4 threads, 2 800 MHz (Turbo 3 400 MHz); GPU NVIDIA GeForce GTX 860 M, 1029 МГц, 640 CUDA cores.
hardware were chosen based on the close price.

MATLAB is chosen as an experimental platform. Hidden layer size varied in the range [4, 56] neurons.

1. The evaluation procedure includes the following steps:
2. Select the hardware (CPU, GPU).
3. Select a value from the range of the hidden layer.
4. Train 10 single neural networks.
5. Evaluate an average execution time.
6. If not all the values of the interval are used, then go to step 3.
7. If not all of the hardware are used, then go to step 2.

The results are shown in Table 4.

As can be seen from the table, there is a 5-fold reduction in training time of a neural network using a GPU with CUDA technology.

3.3 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).

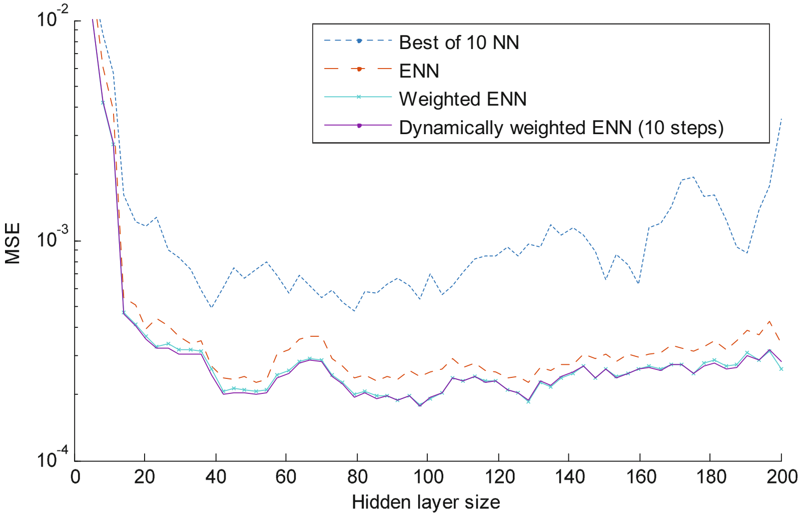

Fig. 2. MSE of different models depending on the size of the hidden layer

Evaluation of different models across testset Dt_set_s1 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.

3.4 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 training time

Hidden layer size	Matlab CPU, s.	Matlab GPU (CUDA), s.	Speed up
4	93,9962	17,2422	5,4515
12	109,2202	19,5071	5,5990
20	86,2531	18,4156	4,6837
28	105,0203	20,0411	5,2402
36	113,4552	18,902	6,0023
44	90,6871	19,8095	4,5780
52	95,4515	16,4389	5,8064
56	107,1488	20,8944	5,1281
Mean			5,3112
Median			5,3459
Minimum			4,5780

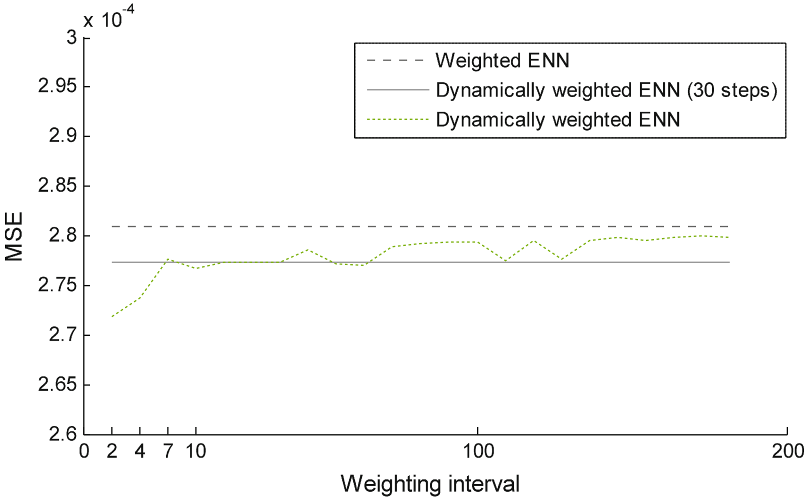

Fig. 3. Evaluation of weighting interval for the dynamically weighted ensemble

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.

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

3.5 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 iteratively trained (without previous access to the data) expert ensemble 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.

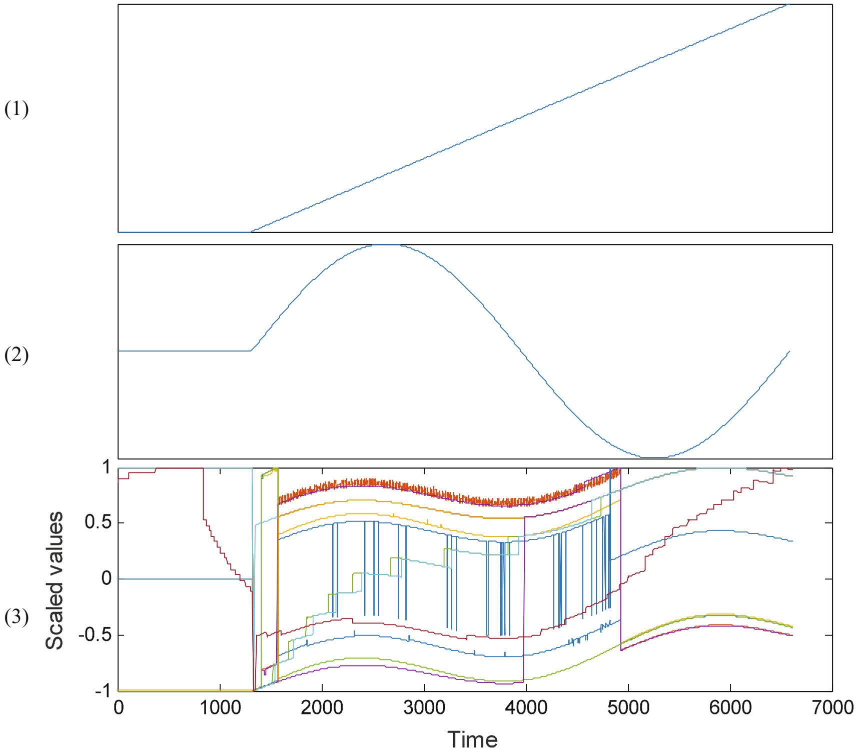

Fig. 4. Modification of Dt_set_s1 data set: (1) Linearly increasing trend; (2) periodic component; (3) a modified multivariate signal

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 dataset are shown in Table 4.

Table 4. Evaluation of NN models for dataset Dt_set_s1

Model	MSE, 10^{-4}	MAE, 10^{-3}
Single neural network	3.51	0.115
Ensemble of neural networks	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

All NN 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 (Table 5).

Table 5. Evaluation of different architectures on a modified set

Model	MSE	MAE
Single neural network	0.6035	0.4843
Ensemble of neural networks	0.0304	0.1274
Weighted ensemble of neural networks	0.0289	0.1212
Dynamically weighted ensemble of neural networks, interval = 10	0.0266	0.1170
Dynamically weighted ensemble of neural networks, interval = 10 with additional training	0.0153	0.0877

4 Conclusion

The developed ENNs significantly reduce the forecasting error in comparison with single NNs.

Dynamically weighted ENN 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 NN 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.

Testing also showed that NN training on a single GPU with CUDA technology gave 5-fold reduction in the time in comparison with the CPU in the specified equipment.

References

1. Quan, H., Srinivasan, D., Khosravi, A.: Short-term load and wind power forecasting using neural network-based prediction intervals. IEEE Trans. Neural Netw. Learn. Syst. **25**(2), 303–315 (2013)
2. Shumway, R.H., Stoffer, D.S.: Time Series Analysis and Its Applications. Springer, New York (2011). doi:10.1007/978-1-4419-7865-3
3. Dalto, M., Matusko, J., Vasak, M.: Deep neural networks for time series prediction with applications in ultra-short-term wind forecasting. In: IEEE International Industrial Technology, Seville, Spain, 17–19 March (2015)
4. Valipour, M., Banihabib, M.E., Reza Behbahani, S.M.: Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J. Hydrol. **476**, 433–441 (2013)
5. Khachumov, V.M., Talalaev, A.A., Fralenko, V.P.: Review of Standards and the concept of monitoring, control and diagnostics of the spacecraft tools building. Softw. Syst. Theor. Appl. **3**(26), 21–43 (2015). (in Russian), Volume 6
6. Emelyanov, Yu.G, Konstantinov, K.A., Pogodin, S.V.: Neural orientation angles and distance of the spacecraft sensor control system. Softw. Syst. Theor. Appl. **1**(1), 45–59 (2010). (in Russian)
7. Marushko, Y.: Using ensembles of neural networks with different scales of input data for the analysis of telemetry data. In: Proceedings of the XV Internship Ph.D. Workshop OWD 2013, Wisla, 19–22 October 2013, pp. 386–391 (2013)
8. Kourentzes, N., Barrow, D.K., Crone, S.F.: Neural network ensemble operators for time series forecasting. Expert Syst. Appl. **41**(9), 4235–4244 (2014). ISSN 0957-4174
9. Elwell, R., Polikar, R.: Incremental learning of variable rate concept drift. In: Benediktsson, J.A., Kittler, J., Roli, F. (eds.) MCS 2009. LNCS, vol. 5519, pp. 142–151. Springer, Heidelberg (2009). doi:10.1007/978-3-642-02326-2_15
10. Parikh, D., Polikar, R.: An ensemble-based incremental learning approach to data. IEEE Trans. Syst. Man Cybern. Part B Cybern. **37**(2), 437–450 (2007)
11. Riedmiller, M.A, Braun, H.: Direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Proceedings of the IEEE International Conference on Neural Networks (ICNN), San Francisco, pp. 586–591 (1993)

Author Query Form

Book ID :	**440434_1_En**
Chapter No.:	**6**

Springer
the language of science

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author
During the process of typesetting your chapter, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query Refs.	Details Required	Author's Response
AQ1	Please confirm if the corresponding author is correctly identified. Amend if necessary.	
AQ2	Please check and confirm if the inserted citation of Table 5 is correct. If not, please suggest an alternate citation.	