Preparing Data and Determining Parameters for a Feedforward Neural Network Used for Short-Term Air Temperature Forecasting

Boris Perelygin [1], Tatiana Tkach [2], Anna Gnatovskaya [3]

[1,2,3] Odessa State Environmental University, Lvivska st., 15, Odessa, 65016, Ukraine

Abstract
This article presents the results of solving the problem of preparing initial data and determining specifications for an artificial feedforward neural network used for short-term forecasting of ambient air temperature values. Based on the requirements for the accuracy of forecasts, the data for network self-learning was optimized, namely, the number of training vectors and their length, the type of the source data itself, the features of creating a training sample from an array of source data were determined. Additionally, the specifications of neural network that provide the required accuracy of forecasts were selected, namely, the requirements for the network neuron activation, and the number of hidden layers.

Keywords
Artificial neural network, short-term forecast, air temperature

1. Introduction

Forecasting is one of the most important tasks in almost all areas of science and life. Predicting weather factors is one of the oldest forecasting tasks because of their great influence on all aspects of human life. Meteorological weather forecasts are a scientifically based assumption about the future state of the weather. The success of modern short-term weather forecasts is quite high, but there are also inaccurate forecasts, especially in cases of abnormal manifestations of the weather. Therefore, research in this area remains relevant at the present time.

In recent decades, along with traditional methods of weather forecasting, the use of artificial neural networks (ANN) for forecasting is considered as a promising area of research [1, 2, 3, 4]. The initial data for weather forecasting for ANN are commonly the results of regular measurements of weather characteristics in the form of numerical values. With the help of ANN, it is possible to model the nonlinear dependence between the future value of a time series, its past values, and the values of external factors [5]. For instance, it is proposed to use fully connected feedforward neural networks to predict the time series of mountain soils humidity [6]. Deep neural networks could be used to predict the meteorological visibility range [7]. Multi-wavelet polymorphic networks could be employed to predict geophysical time series [2]. To predict long-term series, it is proposed to use extreme learning machines [8], convolutional neural networks [9]. In particular, examples of predicting temperature values are presented [1, 3, 10]. However, a comparative analysis of the possibility of using different ANN architectures is carried out in the papers concerned with the aforementioned subject, and, as a rule, the methodology for preparing initial data, numerical estimates of forecasting quality, and the influence of the ANN parameters on these estimates are not sufficiently covered. That means, that these academic papers do not sufficiently cover

solutions problems of selection of parameters of ANN for forecasting meteorological elements. This work aims to fill this gap. Therefore, when implementing this attempt, a well-known feedforward ANN was taken as the subject of research and used for short-term forecasting.

2. Problem statement

The object of the study is the process of using a feedforward ANN to predict air temperature values. The subject of the study is a feedforward ANN designed for short-term forecasting of air temperature values.

In the process of conducting research, it was necessary to find out the impact on the accuracy of forecasts: 1) the parameters of neural network learning data (the length of the training vectors, the number of training vectors, the location of the sample for training in the general series of observations, the types of source data) that provide the best accuracy for short-term forecasts with a lead time of 3 hours, 1 day and 3 days; 2) the influence of the parameters of the neural network on the accuracy of forecasts with the above-mentioned lead time (the number of its hidden layers, the presence of restrictions in the activation functions of neurons of hidden layers).

3. Initial data

The air temperature values were chosen as data for the research because of the continuity of these data and the clarity of the results obtained. The data are a long 15-year series of air temperature values (43569 samples) obtained during regular eight-term observations at the weather station 33837 Odessa from February 01, 2005 to December 31, 2019 (Fig. 1) [11].

Figure 1: A fragment of data on the air temperature of the weather station 33837 Odesa

When forming an array of initial data, the missing air temperature values were interpolated as the arithmetic mean of neighboring temperature values. There were only 6 such omissions in the data, so with a row length of 43569 samples, their correction did not significantly affect the result of the study.

4. Research methodology end tools

Due to the large variability of meteorological quantities in space and in time, the specific value of any value specified in the forecast should be considered as the most likely value that this value will have during the period of the forecast. At the end of the validity period of the short-term forecast, an assessment of its success is made, which is based on the accuracy. Accuracy is the degree of matching, with certain established tolerances, of predictive and actual meteorological values, and phenomena. The accuracy of the temperature forecast is measured alternatively. If the prognostic temperature differed from the actual one by no more than 2.0

°C, then the forecast accuracy is 100%, if the difference is 3.0 °C, the accuracy is 50%, if ≥ 4.0 °C - it's 0% [12] When conducting studies, the accuracy was calculated similarly to the above method, but without an alternative, in the form of the ratio of the number of accurate forecasts (falling within the range of ±2 °C) to the total number of forecasts for a given advance.

In the context of this paper, different types of data should be understood as the actual existing series of observations, the so – called "raw" data, and its two transformations: a centered series – obtained by subtracting the arithmetic mean from all the values of the series; and a normalized series obtained by dividing all the values of the series by the maximum modular value of the series. All series of observations were divided into two large groups: the 1st group of data – for training and the 2nd group of data – for forecasting (Fig. 2).
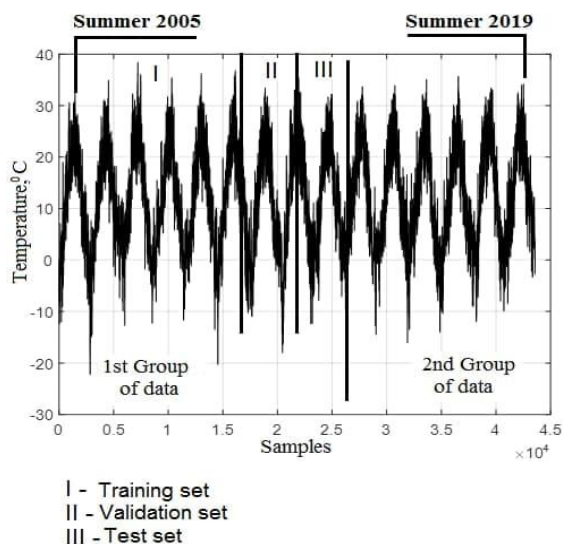

Figure 2: Splitting the source data into groups and arrays necessary for training an artificial neural network and forecasting

The required arrays of initial data were formed as follows from the first group of data intended for training. The whole group was divided into 3 parts. The first part (I in Fig. 2) it was used for training the network (training set). The second part (II in Fig. 2) was used as verification (validation) set to check the quality of training.

Repeated repetition of experiments leads to the fact that the control set begins to play a key role in creating the model, that is, it becomes part of the learning process. This weakens its role, as an independent criterion for the quality of the model – with a large number of experiments, there is a risk of choosing a network that gives a good result using the control set. In order to give the final model proper reliability, the third part of the first group of data was a backup (test) set of observations (III in Fig. 2).

The final model was tested on data from this set to make sure that the results achieved using the training and validation sets are real. According to the obtained data, the indicator of the quality of training was calculated according to the methodology applied to the calculation of the forecast accuracy. During the research, the parameters of the neural network, the length and number of training vectors changed, as well as the place of the beginning of the array changed to assess the seasonal impact of data on the quality of the forecast.

For research, we used a traditional ANN of direct propagation, shown in Fig. 3, the number of inputs of which varied depending on the size of the training set, the number of hidden layers varied from 1 to 3. The neuron activation functions also changed.
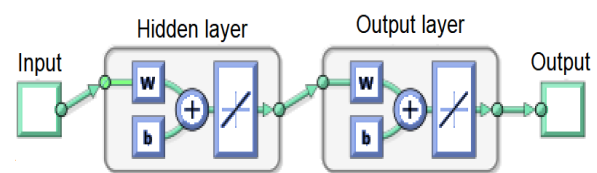

Figure 3: ANN model used in research

The simulation was carried out using a computer with a processor Intel® Core™2 Quad CPU Q8200 2.34 GHz and 12 GB of RAM.

5. Result analysis

When determining the best length of the training vector and the best number of vectors training the neural network in the context of forecasts' accuracy, multiple simulations of the training and forecasting procedure were performed (N cycles – approximately 50). At the same time, the training array size (the length of the vectors and their quantity) was selected in such a way that the training procedure was completed in no more than one hour. Otherwise, the meaning of short-term forecasting with a 3-hour lead time would have been lost, since the result could have been obtained after the forecast time. During multiple simulations, the quality of training and the accuracy of all three forecasts of different timings were evaluated when changing: a) the type of source data ("raw", centered,

normalized), b) when changing the number of hidden layers (1, 2, 3) and c) when changing the activation function from linear to sigmoidal for hidden and output layers. As a result of these studies, 72·N three-dimensional graphs were obtained. Since the volume of this paper does not allow us to present them all, two of them, as an example, are shown in Fig. 4.
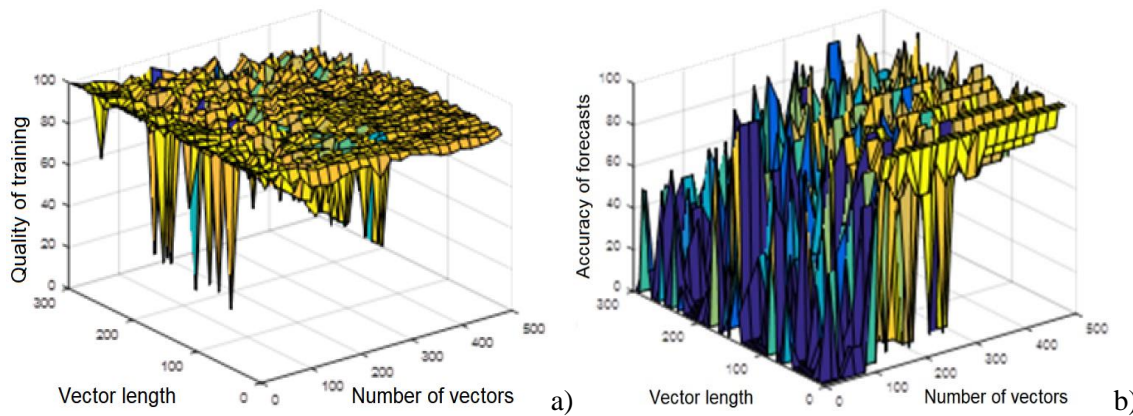

Figure 4: An example of displaying the parameters of the quality of training (a) and the accuracy of forecasts (b)

A number of training vectors is drawn along the abscissa axis, the length of the vectors is drawn along the ordinate axis, and either the value of the training quality or the forecast accuracy parameter for the corresponding advance time is displayed along the application axis. Each point of these graphs was calculated for the specified vector length, the number of vectors, the number of hidden layers of the ANN, the type of activation function and the type of source data. From a mathematical standpoint, the graph in Fig. 4, a shows the quality of the network's approximation of the training array, and in Fig. 4, b – the quality of extrapolation of data on which the network was not trained. The graph in Fig. 4, b clearly shows the instability of forecasts for any length of training vectors and for a small number of them.

The analysis of all the graphs showed that in order to obtain a specific accuracy of forecasts, the initial data for training the network should be presented in the form of 150 vectors with a length of 16 samples each (the circled area in Fig. 5, a), which also provides stable short-term forecasting, one of the results of which is shown in Fig. 6.
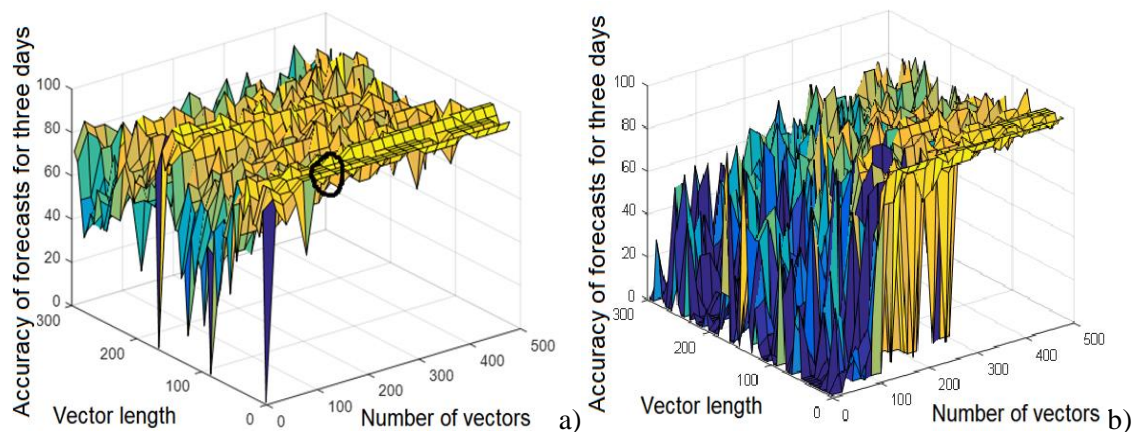

Figure 5: Accuracy of forecasting with a linear function of neuron activation (a) and in the presence of nonlinearity in the activation function of neurons (b) with the same form of initial data and the same number of hidden layers

In addition, the presence of non – linearity (restriction) in the activation function greatly worsens the prediction quality indicator, i.e. accuracy (Fig. 5), and increases the training time by more than 2 times. Therefore, when solving such a problem, the neurons of the network must have a complete linear activation function.
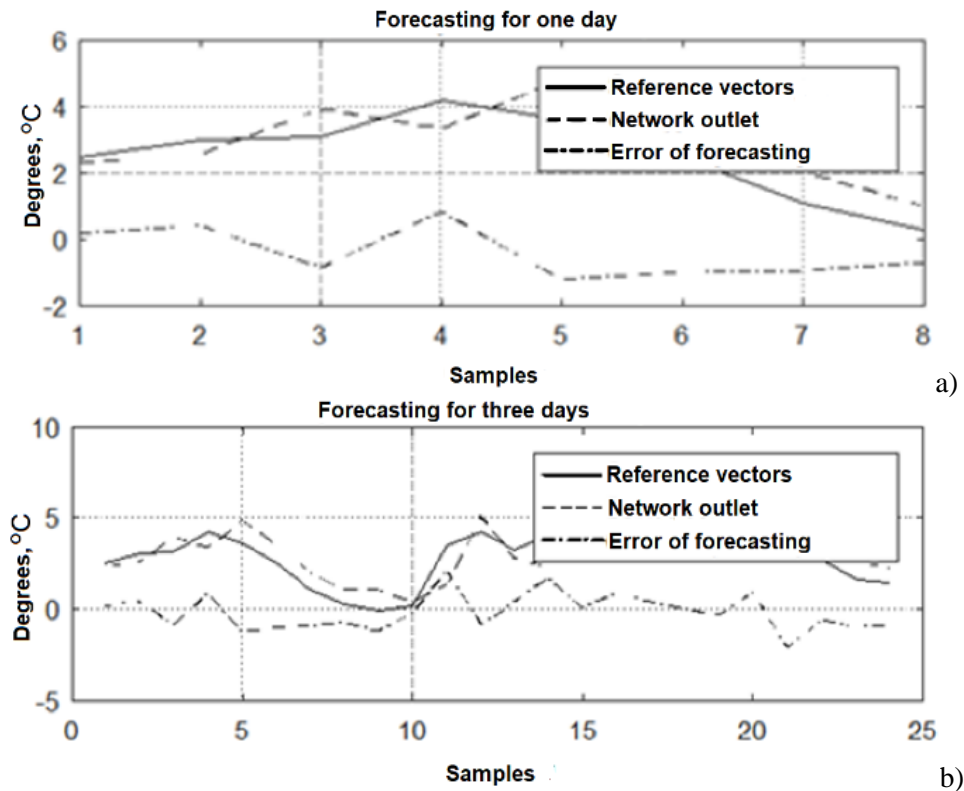
Figure 6: The result of forecasting for one day (a) and for three days (b)

The simulation showed that an increase in the number of hidden network layers does not improve, but also does not worsen the quality of forecasting, the accuracy does not change significantly, but the network architecture becomes more complicated and with the same learning algorithm (the Levenberg-Marquardt algorithm in the error back propagation procedure), the training time increases by a multiple.

The type of source data does not affect the quality of forecasting, the accuracy does not change significantly when replacing the "raw" source data with centered or normalized ones.

6. Conclusions

The analysis of the obtained results made it possible to determine the type of initial data, the volume of initial data, and the main parameters of the feedforward ANN for predicting temperature values:

- the presence of non-linearity (restriction) in the activation function significantly worsens the prediction quality indicator, i.e. accuracy, and increases the training time by more than 2 times, therefore, when solving such a problem,

the neurons of the network must have a complete linear activation function;

- an increase in the number of hidden network layers does not improve, but also does not worsen the quality of forecasting, the accuracy does not change significantly, but the network architecture becomes more complicated and with the same learning algorithm (the Levenberg-Marquardt algorithm in the error back propagation procedure), the training time increases by a multiple;

- the type of source data does not affect the quality of forecasting, the accuracy does not change significantly when replacing the "raw" source data with centered or normalized ones;

- the initial data for training the network should be presented in the form of 150 vectors with a length of 16 samples each, which ensures stable short-term forecasting;

- when training on such a small amount of data, the following condition needs to be observed: when forecasting in a certain season (time of year), earlier data for training needs to be selected, necessarily from exactly the same time of year, otherwise the forecast error increases significantly.

7. References

[1] B. F. Kuznetsov Short-term temperature prediction based on neural networks. // Topical issues of agrarian science. 2019. # 30. Pp. 59-65.

[2] S. N. Verzunov, N. M. Lychenko Multi-wavelength polymorphic network for forecasting geophysical time series // Problems of automation and control. 2017. # 1 (32). Pp. 78-87.

[3] A. S. Kozadaev, A. A. Arzamassev Forecasting of time series using the apparatus of artificial neural networks. Short-term forecast of air temperature. // Bulletin of the Tambov University. Series: Natural and Technical Sciences. 2006. Vol. 11. Issue 3. Pp. 299-304.

[4] A. S. Gribin. Application of artificial neural network algorithms for short-term prognosis: Dis. ... candidate of physical and mathematical sciences. SP-b. 2005. 154 p.

[5] J.W. Taylor, R. Buizza Neural Network Load Forecasting with Weather Ensemble Predictions. // IEEE Trans. on Power Systems, 2002, Vol. 17, Pp. 626-632. DOI: 10.1109/TPWRS.2002.800906

[6] L. I. Velikanova Short-term forecasting of humidity of mountain soils / / Problems of automation and control. – 2015, No. 4. – Pp. 158-166.

[7] S. N. Verzunov. Application of deep neural networks for short-term prediction of visibility range. Problems of automation and control. 2019, #1 (36). Pp. 118-130.

[8] LeiYu, ZhaoDanning, Cai Hongbing Prediction of length-of-day using extreme learning machine // Geodesy and geodynamics. – 2015 V. 6. N. 2. – Pp. 151-159.

[9] Koprinska, Irena et al. Convolutional Neural Networks for Energy Time Series Forecasting // 2018 International Joint Conference on Neural Networks (IJCNN). 2018: Pp. 1-8.

[10] Imran Maqsood, Muhammad Riaz Khan, Ajith Abraham An ensemble of neural networks for weather forecasting Neural Comput & Applic (2004) 13: Pp. 112-122. DOI 10.1007/s00521-004-0413-4

[11] Archive of meteorological data in Odessa. URL:https://rp5.ru/Архив_погоды_в_Одес се (Accessed 19.11.2020)

[12] Guidelines for hydrometeorological forecasting. Ukrainian Hydrometeorological Center. Kyiv, 2019. – 35 p.