Forecast of the near ground air temperature based on the multilayer perceptron model

I V Del and A V Starchenko
National Research Tomsk State University, 36, Lenin Ave., 634050 Tomsk, Russia
E-mail: irina.del@mail.tsu.ru

Abstract. In this study, a multilayer perceptron model is implemented for predicting meteorological values. Based on the known distribution of meteorological values for several previous days, the task was set to predict the values of the near ground air temperature. The overall mean square error for the entire forecast was 3.11 C. Comparison of various optimization methods showed the advantage of the method of Adaptive Moment Estimation. Comparison of the multilayer perceptron model forecasting results with the Weather Research and Forecasting numerical model forecast showed the promise of using neural networks to predict meteorological parameters at weather observation points.

1. Introduction
Dangerous meteorological conditions have a significant impact on the economy and on society. An accurate and timely location-specific forecast is required to avoid financial losses, for example, in agriculture, which is very dependent on meteorological indicators, and the forecast of heavy precipitation, frost, abnormally cold/warm temperatures is of paramount importance for this area. Additionally, dangerous weather events, such as strong and hurricane-force winds, can lead to human casualties. The public can benefit from timely forecasts and warnings of impending dangerous meteorological conditions.

Weather forecasting and modeling of the occurrence of dangerous meteorological conditions are important tasks requiring special attention. To date, predictive models are constantly being improved, their spatial resolution and the accuracy of forecasting various processes are being increased. Thus, in the coming years, the topic of modeling meteorological events will remain relevant.

Forecasting is the most developed field of meteorology. It is worth highlighting the following approaches to scientific weather forecasting: synoptic weather forecasting, numerical and statistical forecasting methods [1, 2], machine learning methods, and neural networks [3, 4].

The machine learning approaches currently in use are clearly data-oriented. Machine learning methods such as linear regression, decision tree regression, LASSO regression, and ridge regression are used for forecasting time series [4].

Also, the following neural network architectures can be distinguished: multilayer perceptron, recurrent neural networks, convolutional neural networks [3, 5].

The aim of this work is to implement and apply the multilayer perceptron model for weather forecasting based on the measured values of the near ground air temperature. According to the known distribution of meteorological values for the previous few days, the task is to forecast the values of the near ground air temperature.
2. Materials and Methods

2.1. Creating a multilayer perceptron model

The developed artificial neural network (ANN) model is based on one of the neural network architectures which is called the multilayer perceptron (MLP) model. A multilayer perceptron is a class of feedforward artificial neural networks which consists of at least three layers: input, hidden, and output [3]. One of the most important aspects of a deep neural network is its activation function. The activation function determines the output value of the neuron depending on the result of the weighted sum of inputs and the threshold value [3].

While going through MLP training, supervised learning and the backward propagation of errors algorithm are used. A three-level structure was chosen (input layer, one hidden layer, and output layer) with the Rectified Linear Unit (ReLU) activation function for the hidden one and without any activation function for the output layer since we are interested in numerical values without any transformation.

The input and hidden layers contain 48 neurons. The output layer contains 24 neurons.

The initial data for the study were obtained in the Atmosphere JUC of the Institute of Atmospheric Optics SB RAS for the period from 1 January, 2020 to 30 October, 2020. The dataset size is 6336 hourly observations.

The task is to forecast the near ground air temperature for 24 hours using the hourly temperature for the past two days (48 hours).

After the data are collected, the data preprocessing procedure is performed (restoring omissions). The missing data are recovered using linear interpolation.

The selected data were split into two samples: a training sample corresponding to 80% of the main sample, and a test sample corresponding to 20% so that the network's forecasting ability could be tested after the training phase.

The network was trained for a fixed number of epochs (500). The number of epochs shows how many times the model is exposed to training. An epoch is a pass forward or backward for all training examples. The number of epochs was chosen experimentally.

2.1.1. ReLU activation function. The Rectified Linear Unit is the most commonly used activation function in deep learning. The ReLU function returns 0 if it takes a negative argument, but in the case of a positive argument, the function returns the function argument itself.

2.1.2. Optimization algorithm. This neural network is optimized by the method of Adaptive Moment Estimation (a variant of stochastic gradient descent) [7]. The rule for updating the weights is determined by using the estimates of two different moments:

\[ m_N = \alpha_1 m_{N-1} + (1 - \alpha_1) h_N, \quad v_N = \alpha_2 v_{N-1} + (1 - \alpha_2) h_N^2, \]

The first uses the previously calculated values of partial derivatives, and the second uses their squares. The method of Adaptive Moment Estimation (Adam) is considered resistant to the selection of hyper-parameter values \( \alpha_1, \alpha_2 \). The calculated moments are corrected using the formulas

\[ m_N = \frac{m_N}{1 - \alpha_1^N}, \quad v_N = \frac{v_N}{1 - \alpha_2^N}, \]

and then the weights are recalculated according to the formula

\[ w_{N+1} = w_N - \frac{\eta}{\sqrt{v_N + \varepsilon}} m_N. \]

The entered designations:

\[ h_N = \nabla_{w_N} F(w_N), \]
where $F$ – the target error function which depends on the parameter; $w$ – the weight coefficients of the neural network; $\varepsilon$ – the smoothing factor which allows avoiding division by 0.

2.1.3. Metrics. The metric for the problem under consideration in this paper will be the Root Mean Square Error (RMSE) calculated for each forecasted time step (from 4 September, 2020 to 30 October, 2020).

The mean squared error (MSE) is taken as the loss function since it corresponds to the scale of the previously accepted metric (RMSE):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - d_i)^2,$$

where $y_i$ is the predicted output value and $d_i$ is the real output value. The loss function is used to calculate the error between the actual and received values. The overarching aim is to minimize this error [6].

3. Results and discussion
The degree of proximity of the graphs in Figure 1 indicates a satisfactory network setup.

The overall RMSE value for the entire forecast was 3.11 $^\circ$C.

Table 1 shows the results of the RMSE values for some days from the test sample. When comparing these values with each other, the ANN appears to be worse at forecasting on some days, and on others, it is better.

| Date       | 04.09.20 | 15.09.20 | 16.09.20 | 18.09.20 | 29.09.20 | 30.09.20 | 08.10.20 | 11.10.20 | 20.10.20 |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| RMSE       | 5.95     | 1.36     | 1.08     | 3.87     | 6.22     | 0.78     | 2.68     | 0.59     | 2.42     |

For example, Figure 2 shows a graph comparing the forecasted hourly temperature with the observations for 15 September, 2020 (RMSE = 1.36).
3.1. Comparison of the ANN forecasting results with the Weather Research and Forecasting numerical model forecast

The Weather Research and Forecasting (WRF) is a numerical weather forecast model designed for both atmospheric research and operational forecasting [8].

WRF is a set of programs implementing the model calculation. WRF supports two types of dynamic solvers, a data assimilation system, and software components for parallel computing and system extensibility. The WRF model has a wide range of applications in meteorology and is used on the scales ranging from meters to thousands of kilometres.

It can be concluded from Figure 3 that the hourly temperature forecast for 15 September, 2020 obtained using the multilayer perceptron model is more accurate than the forecast of the numerical WRF model.
3.2. Variation analysis of a multilayer perceptron model

The following gradient optimization methods were applied: Adaptive Gradient method (AdaGrad), Root Mean Square Propagation method (RMSProp), Adam, modification of Adam (Adamax), Nesterov–accelerated Adaptive Moment Estimation method (Nadam), Adadelta is another improvement of AdaGrad which aims to reduce the monotonically decreasing learning rate [7].

These methods enable the finding of an approximate solution in an acceptable time when working with big data. As a rule, methods have many parameters which need to be selected, and these parameters may vary for different tasks.

Follow The Regulated Leader (FTRL) optimization method was also applied.

Comparing the RMSE values for different optimization methods and the size of the input data (24, 48, 72, and 96 observations), it can be concluded that the minimum RMSE value is given by the Adam optimization method. However, all other optimization methods, except Adadelta, give a satisfactory result, as can be seen from Figure 4.

![Figure 4. Comparison of the RMSE values for different input data sizes and different optimization methods.](image)

4. Conclusions

In this paper, a model of a multilayer perceptron was constructed and applied to forecast the near ground air temperature in the city of Tomsk.

A comparative analysis of real data and the results obtained using the multilayer perceptron model indicate that the neural network forecasting method can compete with standard forecasting methods.

Acknowledgments

The problem was stated and the mathematical model was formulated by A. Starchenko, and this work was supported by the Ministry of Science and Higher Education of the Russian Federation under project of Regional Scientific and Educational Mathematical Center of Tomsk State University.
References

[1] Marchuk G 2012 *Numerical Methods in Weather Prediction* (Elsevier) p 288
[2] Starchenko A V, Shelmina E A and Kizhner L I 2020 *Atmosphere* 11 1148
[3] Haykin S 2008 *Neural networks, a comprehensive foundation* (New York: Macmillan College Publishing Company) p 1104
[4] Raschka S 2016 *Python Machine Learning* (UK: Packt Publishing) p 340
[5] Ren X, Li X, Ren K, Song J, Xu Z, Deng K and Wang X 2021 *Big Data Research* 23 100178
[6] Chollet F 2017 *Deep Learning with Python* (New York: Manning Publications Company) p 384
[7] Dogo E M, Afolabi O J, Nwulu N I, Twala B and Aigbavboa C O 2018 2018 *International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* pp 92-99 doi: 10.1109/CTEMS.2018.8769211
[8] Powers J G, Klemp J B, Skamarock W C, Davis C A, Dudhia J, Gill D O, Coen J L, Gochis D J, Ahmadov R, Peckham S E and et al 2017 *Bull. Am. Meteor. Soc.* 98 1717