Creating an artificial neural network for predicting the dynamics of retrospective yield series

A F Rogachev\(^1,2\)

\(^1\)Volgograd State Agricultural University, 26, Universitetskii Avenue, Volgograd, 400002, Russian Federation
\(^2\)Volgograd State Technical University, 28, Lenina Avenue, Volgograd, 400005, Russian Federation

E-mail: rafr@mail.ru

Abstract. The article considers the specifics of constructing an artificial neural network (ANN) for predicting the dynamics of time series levels (TS) of grain yield in arid conditions on the example of the Volgograd region. In order to justify the choice of the architecture and hyperparameters of ANN, a preliminary statistical analysis of the simulated TS was performed. The autocorrelation functions of the retrospective yield series constructed in this case can be taken into account when choosing ANN hyperparameters for predicting grain yield. A number of ANN architectures based on recurrent layers, including LSTM, were analyzed. The best results of neural network modeling are obtained for cascading groups of layers of a series-parallel architecture. The proposed neural network technology for predicting TS yield levels using a pre-forecast autocorrelation analysis of retrospective levels reduces the error of short-term forecasting of grain yield in the arid natural and climatic conditions of the Lower Volga region. Taking into account the results of autocorrelation analysis allows you to choose the values of ANN hyperparameters more reasonably. The achieved accuracy of the regression problem was 82.87%, which can be considered acceptable for planning agricultural production for 1-2 years. The ways of improving the accuracy of the neural network solution of the problem of predicting productivity in the arid conditions of the Lower Volga region using retrospective TS of productivity are formulated.

1. Introduction

Agricultural production is characterized by a significant dependence on hydrothermal conditions, especially in cases of insufficient moisture, typical, for example, for the Lower Volga region of Russia. For the management of agricultural production, an accurate forecast of the yield of various crops is important. A number of methods for predicting TS yield are known [1, 3, 5, 6, 7, 10], however, their accuracy is insufficient.

The cultivation of agricultural crops in conditions of insufficient moisture supply is characterized by a high coefficient of variation of time series (TS) of yield levels, amounting to more than 0.31 for a group of cereals, which limits the use of multi-factor trend and seasonal models [1, 9]. The mentioned specificity of TS yield causes significant errors in the assessment of risks taken into
account when planning agricultural production using approaches and methods of nonlinear dynamics [1-3].

Effective approaches include cellular automata, cognitive modeling based on fuzzy cognitive maps [21], artificial intelligence (AI) methods, including deep machine learning [20], modeling based on artificial neural networks (ANN) [21]. Neural network modeling technologies based on artificial neural networks allow us to identify hidden internal patterns that determine the dynamics of the simulated processes, and to adequately model interannual fluctuations in productivity. The use of modern methods of artificial intelligence [3], in particular, ML and neural network, allows to increase the reliability of the obtained forecasts.

The purpose of the study is to increase the reliability of the results of neural network modeling for predicting yield by taking into account the internal patterns of the dynamics of its changes in previous years. For this purpose, it is advisable to use the results of autocorrelation statistical analysis of retrospective TS yields of the studied crops.

2. Methods

When conducting neural network modeling, TS of grain crop yields in general were used according to the data of the State Federal Statistics Service for the Volgograd Region, formed for the period 1950-2018. To identify the cyclic components of TS, the values of the autocorrelation coefficients were calculated according to the dependence (1).

\[
\gamma_k = \frac{\sum_{l=1}^{n-k} x_{t+l} x_{t+l-k} - \sum_{l=1}^{n-k} x_{t+l} \sum_{l=1}^{n-k} x_{t+l-k}}{\sqrt{\sum_{l=1}^{n-k} x_{t+l}^2 - \left(\sum_{l=1}^{n-k} x_{t+l}\right)^2/(n-k)} \sqrt{\sum_{l=1}^{n-k} x_{t+l-k}^2 - \left(\sum_{l=1}^{n-k} x_{t+l-k}\right)^2/(n-k)}}
\]

where \( n \) is the sample size.

The preliminary statistical analysis, taking into account the specifics of the arid conditions of the Volgograd region, provided for testing the hypothesis that the TS yield corresponds to the normal law according to the statistical criteria of Pearson and Kolmogorov-Smirnov. The ANN were built using the Python Keras libraries in the Google Colaboratory cloud computing environment. This approach simplifies the development of hybrid neural network architectures, including their parameterization and modeling of the TS under study.

3. Results and discussion

3.1 Preliminary statistical analysis of time series

For the sample population of grain yields as a whole, the calculated value of the Pearson criterion exceeded the value of 51, which significantly exceeds the critical value. component 9.5, determined from the table at (significance level \( \alpha = 0.05 \) at \( v = 4 \)).

The analysis of the retrospective yield data showed that the empirical distribution of the retrospective yield levels TS of grain crops differs significantly from the normal one, which is confirmed by each of the mentioned consent criteria.

In order to substantiate the architecture and macroparameters of mathematical models of grain yield in arid conditions, which must be taken into account when constructing the ANN, and to increase the reliability of the obtained forecasts, an autocorrelation analysis of the studied BP was carried out.

The results of the autocorrelation analysis for the group "cereals in general" are shown in Figure 1. The nature of autocorrelation functions performed for different groups of grain crops calculated according to (1) demonstrates the presence of statistically significant cyclic components. Hence, the subsequent neural network modeling of crop yields must be performed taking into account the obtained characteristics of their cyclicity.
The diagrams of autocorrelation functions show statistically significant peaks corresponding to lags of 1, 2, 3, 4, and twelve-year duration (Figure 1). The diagram shows that the most pronounced "peaks" are at lags of three and twelve years. A possible hypothesis to explain the 12th cycle may be the superposition and multiplicity of 2 and 3 cycles. Therefore, the cyclical nature of grain yield, confirmed in the conducted studies, as an endogenous attribute of the nature of dynamics, must be taken into account when constructing neural network forecasts of yield.

Thus, the use of pre-forecast statistical analysis, the methods of which are implemented by means of built-in functions in many computer programs, is appropriate as a means of increasing the reliability of forecasts.

3.2 Building artificial neural networks

The main tasks of the preliminary statistical analysis were to justify the numerical values of the hyperparameters of neural networks in the formation of their architecture even before the start of the training procedure. In particular, the size of the "window" when training a neural network was taken 10..14 years. The formation of the training and test samples was made from the values of the original TS, provided for the augmentation of the original data and the complete absence of the intersection of the training and test samples. We tested variants of networks with recurrent layers of various types, both recurrent and convolutional.

The practical implementation of the developed ANN models was carried out by means of Google Collaboration. The training of the constructed ANNs was carried out in an automated mode using the built-in Keras tools using the Adam optimizer [8].

An example of a sequential architecture network based on a recurrent LSTM layer with 24 neurons and a fully connected "Dense" layer with 4 neurons is shown in Figure 2.
Taking into account the limited TS levels and to prevent the phenomenon of ANN overfitting, "Dropout" layers and other normalization and regularization options were provided before the output layer.

However, the results of using such a network, as well as other architectures with recurrent layers for predicting yield, were not adequate enough.

In order to improve the quality of neural network modeling, various networks of combined architectures were analyzed. The best results were shown by architectures based on cascading series-parallel structures, which provide for the separation of information flows and subsequent concatenation of the outputs of recurrent and convolutional cascades. The duration of training of such "heavy" architectures was about an order of magnitude higher than previously studied options, but the quality of prediction increased by several percent.
Figure 3. ANN structure of the combined architecture for yield modeling and forecasting

Neural network analysis of TS data on the yield of various grain crops, including the group "Grain crops in general", with multiple random runs of ANN, the achieved accuracy was 82...87%, which can be considered an acceptable result.

4. Discussion

For the TS yield of various cereals of the crop group, including the group "Cereals in general", with repeated runs of ANN, the achieved accuracy was 82...87%, which, when compared with other methods, can be considered a relatively acceptable result. The main problem with further improving accuracy is the limited length of the initial yield TS.

To solve this problem, the following approaches are possible.

It seems promising to simultaneously analyze TS yields, both for a set of biologically similar agricultural crops, for example, cereals, and to generate data from geographically similar or different regions [2]. At the same time, the data sets for ANN training are significantly expanded, which can positively affect the quality of network training and the accuracy of forecasting.

The second direction of increasing accuracy is to improve the architecture and hyperparameters of ANN, including the use of automated parameter optimization tools. Since such programs require significant computational resources and time, it is advisable to use tools that increase the speed of training neural networks.

The third direction is related to the implementation of the methods of augmentation of the initial TS data, however, in relation to the neural network solution of the regression problem, this approach requires additional research.

A separate direction is to supplement the numerical retrospective yield data supplied to the input of the neural network with additional information containing data of a different physical nature. In relation to the creation of a neural network control system for programmable agricultural production, such data can be used as graphical images obtained by remote sensing methods of agricultural crops, for example, obtained for several characteristic periods of their vegetation. Before feeding to the output layer of the neural network, it is necessary to perform concatenation of various information flows with possible normalization and regularization.

5. Conclusion
The proposed neural network technology for predicting TS yield levels using a pre-forecast autocorrelation analysis of retrospective levels reduces the error of short-term forecasting of grain yield in the arid natural and climatic conditions of the Lower Volga region.

Taking into account the results of autocorrelation analysis allows you to choose the values of ANN hyperparameters more reasonably. The achieved value of the accuracy of solving the regression problem was 82.87%, which can be considered a relatively acceptable result for planning agricultural production for 1-2 years.

The ways of increasing the accuracy in relation to the neural network problem of predicting yield in the arid conditions of the Lower Volga region using retrospective TS of yield are formulated.

6. Acknowledgments

The publication has been prepared with the financial support of the RFBR and the Administration of the Volgograd region under project No. 19-416-340014.

References

[1] Ilyasov R, Yakovenko V, Malyutina T, Tkachenko D and Magomadov E 2019 Methods Of "New Econometrics" In Study Of Interrelations With Variable Structure European Proceedings of Social and Behavioural Sciences 58 497-504
[2] K E Tokarev et al 2020 IOP Conf. Ser.: Earth Environ. Sci. 421 042015
[3] Popova E V, Kumratova A M, Nedogonova T A and Vasilenko A I 2018 Identification of trend-seasonal components in the time series of economic processes Modern economy: problems and solutions 6(102) 20-30
[4] Rogachev A 2015 Economic and Mathematical Modeling of Food Security Level in View of Import Substitution Asian Social Science 11(20) 178-185
[5] Rogachev A F 2018 Methodological approaches to obtaining and processing of remote sensing data for justification of reclamation activities
[6] Yurchenko I F 2018 Information support for decision making on dispatching control of water distribution in irrigation Journal of Physics: Conference Series 1015(4) 042063
[7] Khubaev G N 2010 On the influence of errors of independent variables on the choice of the composition of factors and the structure of the regression equation Management of economic systems: electronic scientific journal 23 243-249
[8] Melikhova E V and Rogachev A F 2019 Studies in Computational Intelligence 826 1193-1201
[9] Gooijr J G and Hyndman R J 2005 25 Years of IIF Time Series Forecasting: A Selective Review. Tinbergen Institute Discussion Paper 68(4)
[10] Huang B Q, Rashid T and Kechari M-T 2006 Multi-Context Recurrent Neural Network for Time Series Applications International Journal of Computational Intelligence 3(1) 1304-1386
[11] Zahng G, Patuwo B E and Hu M J 1998 Forecasting with Artificial Networks: The State of the Art International Journal of Forecasting 14 35-62
[12] Hraskoa R, Pacheco A and Krohling 2015 Procedia Computer Science 55 990-999
[13] Breeders N, Spsihilova N and Taspaev S 2015 The use of neural network technologies in predicting the efficiency of grain production Bulletin of the Orenburg State Agrarian University 1(51) 216-219
[14] Yun G and Medinsky D 2017 The use of unmanned aerial vehicles in agriculture Science and technology 36(4) 35-41
[15] Lozovoy Ya S and Sekirin A I 2019 Solving the forecasting problem using neural networks).
[16] Rogachev A and Melikhova E 2020 Creating a neural network system for forecasting and managing agricultural production using autocorrelation functions of time series E3S Web of Conferences 164 06005
[17] Smetanin A A, Goncharov P V and Ososkov G A 2020 The choice of deep learning methods for solving the problem of recognizing plant diseases in a small training sample System analysis in science and education 1 30-38