UDC 004.853
https://doi.org/10.23947/2541-9129-2019-3-17-22

FORECASTING FOREST FIRE BURNING AREA USING MACHINE TRAINING

Filippenko V. A., Zotov A. V.
Don State Technical University, Rostov-on-Don, Russian Federation

philippenko@inbox.ru,
azotzot@gmail.com

The objective of this article is to create and train an artificial neural network based on a data set containing various climatic parameters and future fire area as an output parameter that the authors intend to predict. Such a “set” of data is usually available for research and study. Before training the neural network model, the data set is divided into two samples: a sample for training, which is about 90% of the set; and a sample for testing the trained model. In setting the task, the authors select and analyze the known data on the fires that occurred in Montesinho Park, compare the models trained on these data with and without normalization. As a result, two examples are given of a qualitative demonstration of graphs of absolute error changes of fire areas, which are projected using the created and trained model.

Keywords: burning area, machine training, model, neural networks, *Keras*, forecasting, forest fire.

УДК 004.853
https://doi.org/10.23947/2541-9129-2019-3-17-22

ПРОГНОЗИРОВАНИЕ ПЛОЩАДИ ГОРЕНИЯ ЛЕСНОГО ПОЖАРА С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

Филиппенко В. А. , Зотов А. В.
Донской государственный технический университет, Ростов-на-Дону,
Российская Федерация
philippenko@inbox.ru,
azotzot@gmail.com

Целью настоящей работы является создание и обучение искусственной нейронной сети на основе набора данных, содержащих различные климатические параметры и будущую площадь пожара в качестве выходного прогнозируемого параметра. Такой набор данных является, как правило, доступным для исследования и изучения. Перед обучением модели нейронной сети набор данных разделяют на две выборки — выборка для обучения, которая составляет около 90 % от набора, и выборка для тестирования обученной модели. В постановке задачи авторы выбирают и анализируют известные данные о пожарах в парке Монтезиньо (*Montesinho*), сравнивают модели, обученные на этих данных с нормализацией и без нее. В качестве результата приведены два примера графиков изменения абсолютной ошибки площадей пожара, прогнозируемых с помощью созданной и обученной модели.

Ключевые слова: площадь горения, машинное обучение, модель, нейронные сети, *Keras*, прогнозирование, лесной пожар.

Introduction. A forest fire is a natural and uncontrolled spread of fire over forest areas. According to the Federal Forestry Agency in a week from June 3 to June 9, 2019 in 45 regions of Russia, forest fire forces and contractors extinguished 354 forest fires on the area of 5783.2 hectares, including 98 fires on the area of 1790.05 hectares, which were extinguished during the weekend of June 8-9. Due to smoke in fires, about 300 thousand people die every year. As a result of the combustion of biomass, an aerosol-gas mixture is formed, which represents an ecological and toxicological risk for humans.

Fire-fighting service personnel should be provided with the most effective fire-fighting equipment and equipment for natural phenomena elimination. However, often this is not enough to fight this dangerous phenomenon effectively. Strategic planning and resource allocation, such as the provision of a sufficient number of fire-fighting aircrafts or ground crews, can significantly improve the chances of fire control. But you need to calculate the amount of resources that can take a lot of time.

One way to solve this problem can be the use of neural networks. In the present work, the authors used the data on fires in the Montesinho Park in Portugal to train and test the neural network. This set of data is available for research and work [1]. The authors use the *Keras* neural network library [2], written in the *Python* programming language [3, 4].

Data preparation. The Montesinho Park fire data were chosen as a training material for the neural network model due to the fact that the complex fire hazard indicator of V. G. Nesterov used in the Russian Federation contains fewer parameters, and this may be the cause of lower results in the training of the model. The set of parameters used by the authors, in addition to traditional ones, contains the following parameters: moisture content of forest litter and soil, flame characteristics, anthropogenic factor and thunderstorm activity. The following are additional parameters of the rating system of forest fire danger, which were used in the formation of this set [5]:

• probability of fire (*Fine Fuel Moisture Code, FFMC*);
• coal moisture rate (*Duff Moisture Code, DMC*);
• drought rate (*Drought Code, DC*);
• index of the initial spread system (*Initial Spread Index, ISI*).

All meteorological data for the calculation of the above mentioned components can be requested from the nearest meteorological service. Since this data set contains quite a lot of climatic parameters, with the help of the created and trained model it will be possible to predict the future area of a fire not only for the Montesinho Park, but also for any other similar territory.

Complete data in the set:

• *X* — X-axis spatial coordinate on the Montesinho Park map: from 1 to 9;
• *Y* — Y-axis spatial coordinate on the Montesinho Park map: from 2 to 9;
• "month" — month of the year: from January to December;
• "day" — day of the week: Monday to Sunday
• *FFMC* — the index of ease of ignition of the fuel from the *FWI* system over the interval 18.7–96.2;
• *DMC* — the index of coal moisture content rate from the *FWI* system over the interval 1.1 to 291.3;
• *DC* — the index of drought rate from the *FWI* system over the interval 7.9–860.6;
• *ISI* — the index of initial distribution from the *FWI* system over the interval from zero to 56.1;
• "Temperature" — temperature in the range of 2.2–33.3°C;
• relative humidity from 15.0 to 100 %;
• "Wind" — wind speed from 0.4 to 9.4 km/h;
• outside rain from 0.0 to 6.4 mm/m^2;
• "Area" — burned forest area from 0.00 to 1090.84 ha.

All the parameters in the set are changed in different ranges. In order to improve the prediction accuracy of the model, it is necessary to normalize the data. One way to normalize the data is to subtract the mean from each parameter and divide it by the standard deviation. After these actions, the average value will be zero and the variance will be one. In this case, the data in each column will vary from -1 to +1, but with this method of normalization, some columns may have negative values, which may not be the case for some parameters. You can use the *MinMaxScaler().fit_transform()* procedure to resolve this problem [6], which converts all data to the range 0 ... +1. This model is trained by "supervised training". In this case, the data is divided into two parts — the data for training and the correct answers for this data. The data for training are needed to train the model, and the answers are needed to recalculate the weights on the edges of the neural network graph when the predicted value and the actual value do not match. Before training, we will randomly divide this data into a training sample and a test sample. The training sample is part of the dataset used to train the model. It will be about 90 % of the set. The test model is 10 % of the dataset and is used to test the effectiveness of the model. The test data will not participate in the training of the model, it is used only to verify the functionality. In the following, such modeling can be associated with the classical construction of models and the calculation of technosphere safety indicators [7, 8].

Creation of a model. In the testing of different models for the dataset under consideration, the model with 6 layers was the best: the input layer with 24 neurons, 4 hidden layers containing 48, 96, 48, 24 layers, and the output neuron.

The following activation functions were used:

• linear — on the first, second and fifth layers;

• sigmoid — on the second and third layers, it allows you to amplify weak signals without being saturated with strong ones;

• *selu* — on the output neuron, increases the convergence rate of the neural network.

When compiling the model, "*adadelta*" was used as a gradient descent type optimizer. *Adadelta* updates smaller weights that are too frequently updated, but, in contrast to *Adagrad*, instead of the full amount of updates will use the average value with respect to the history of the square of the gradient.

As an error function, which will be used by the optimizer in the error back propagation algorithm, we choose the standard error, as a metric- "*mae*", the average absolute error.

Training. In the 500-stage of training, the average absolute error is 4.6, so the model in the predictions will be wrong in general by 4.6 hectares, which the authors consider satisfactory. Fig. 1 shows the curve of the error change.

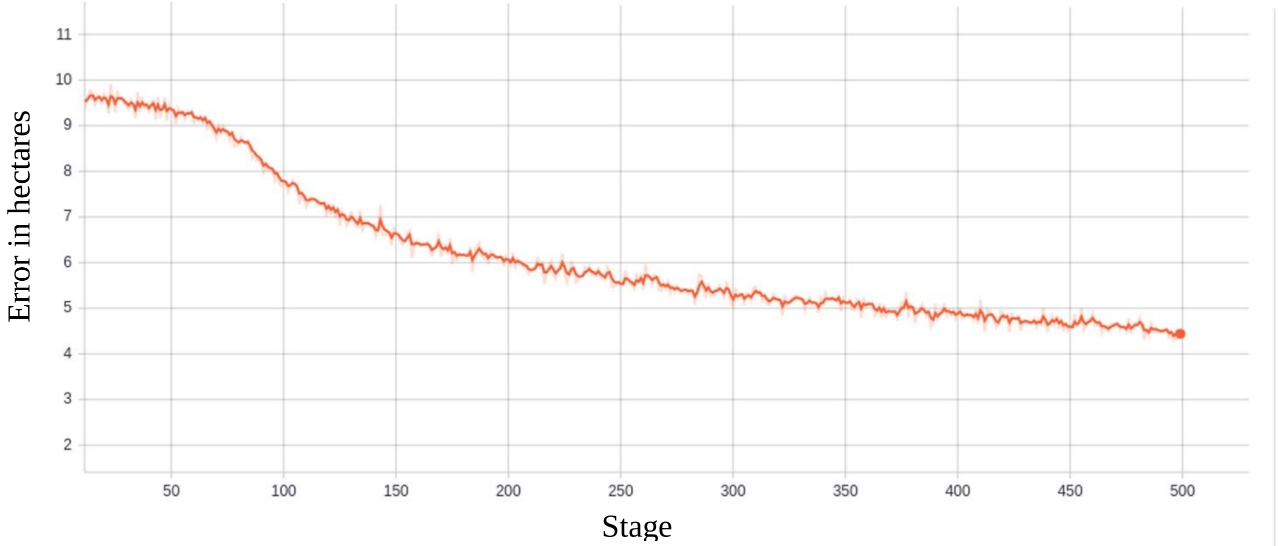

Fig. 1. Curve of an absolute error change at training on data with normalization

Fig. 2 shows a graph of an absolute error change with unnormalized data, which proves that the normalized data is better than the initial data.

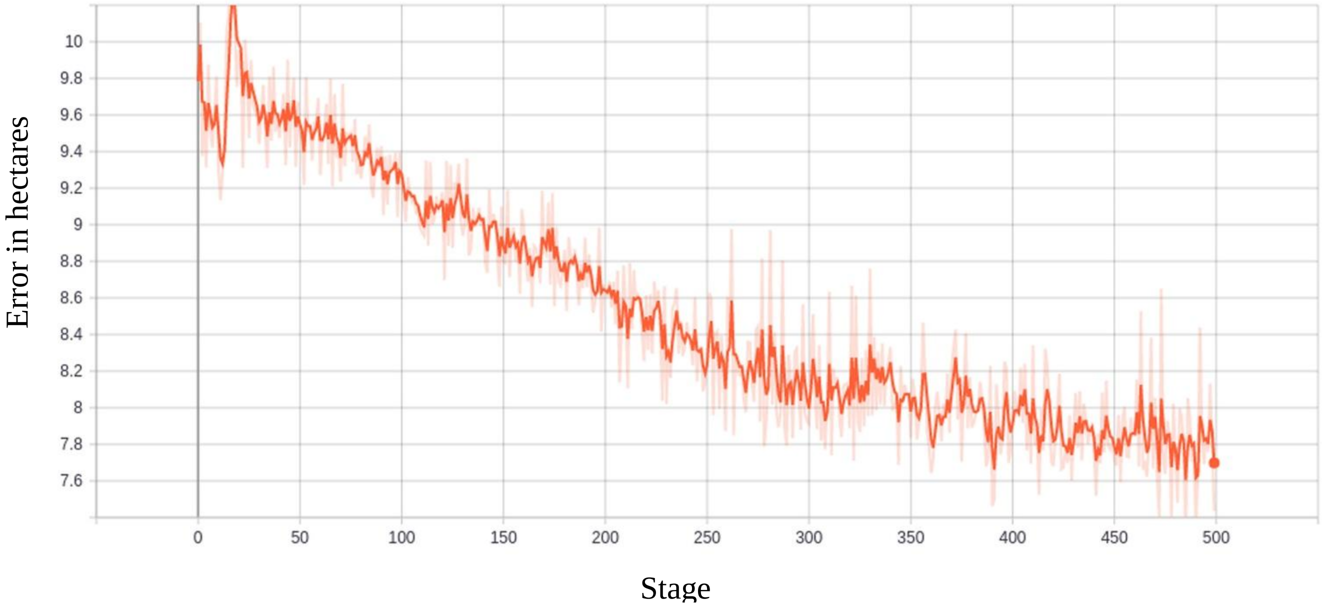

Fig. 2. Curve of an absolute error change at training on data without normalization

Forecasting. Fig. 3 provides a graph illustrating the results of the neural network operation. On the graph, the orange line is the actual burning area; the blue line is the predicted area. As it can be seen in the figure, the dynamics of the curves for each record is almost identical, which shows the good performance of the trained model in forecasting.

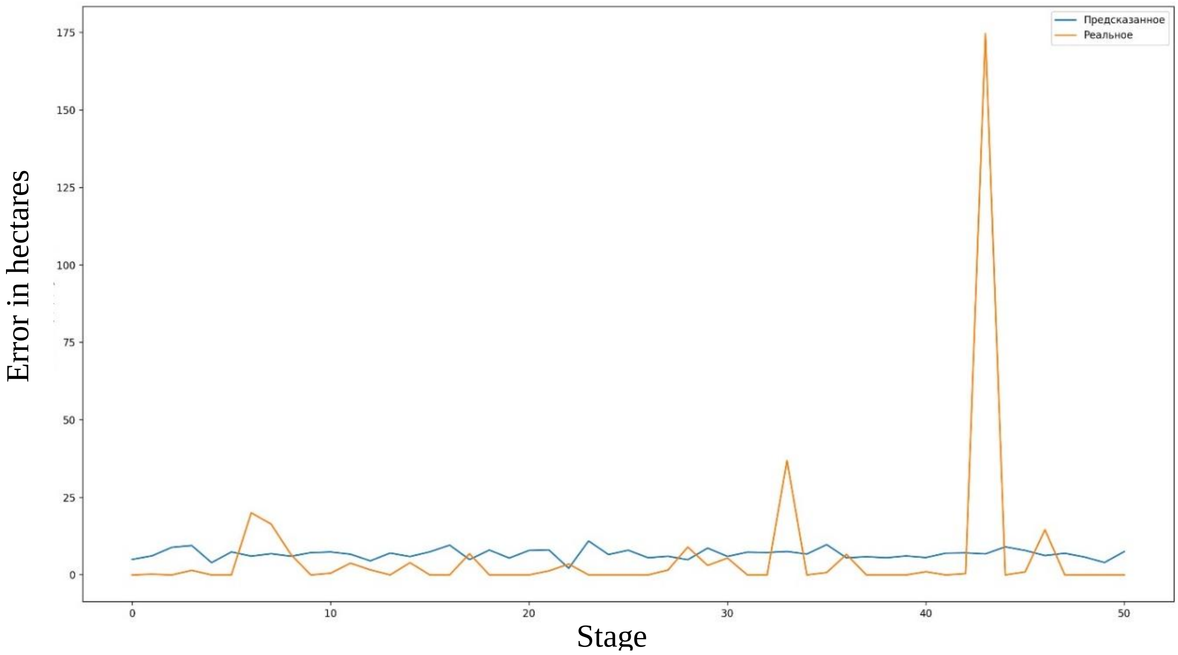

Fig. 3. Forecasting using the trained model with the normalized data

Fig. 4 demonstrates the forecasting graph of another model, which was trained with the help of unnormalized data. Obviously, the first model is more effective.

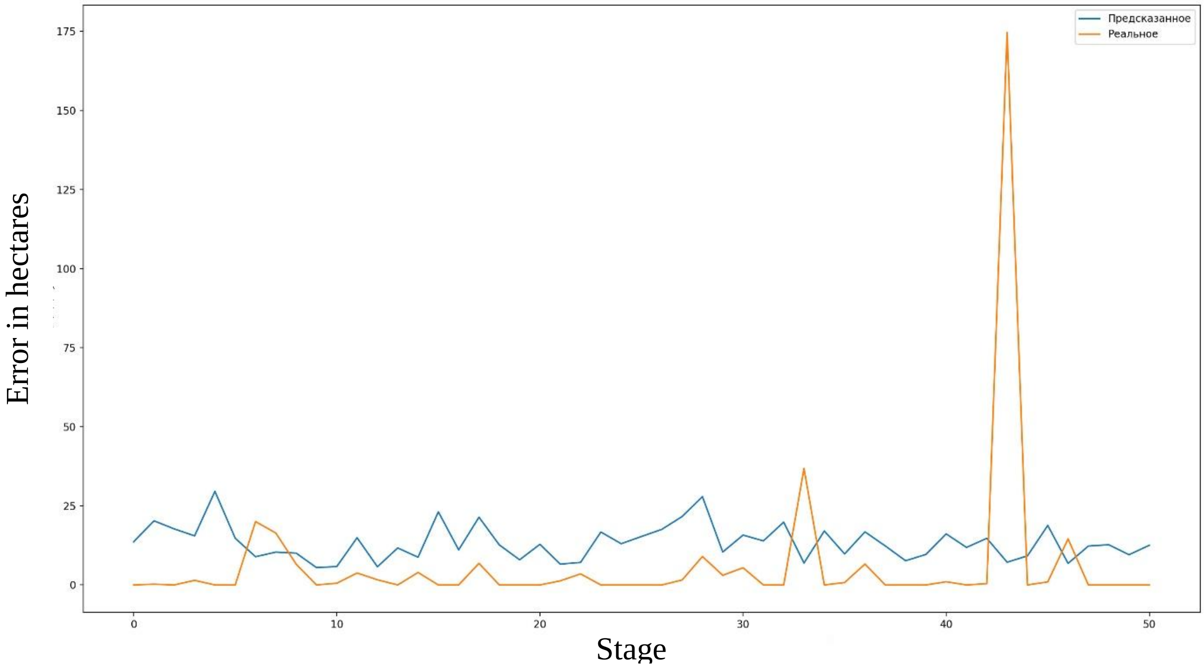

Fig. 4. Forecasting using the trained model with unnormalized data

Conclusion. In this paper, a model of an artificial neural network was created and trained on a set of data containing different climatic parameters and the future fire area in hectares. This area is the output parameter that the authors are going to forecast. As a rule, this dataset is available for research and study. Before training the neural network model, the dataset was divided into two samples: a sample for training, which is about 90 % of the set, and a sample for testing the trained model. In the formulation of the problem, the authors choose and analyze the known data on the fires that occurred in the Montesinho Park, compare the models trained on these data with and without normalization. As a result, two examples of demonstration of the absolute error graphs of the fire areas predicted by the created and trained model are given.

References

1. Cortez, P.A. Morais, A. Data Mining Approach to Predict Forest Fires Using Meteorological Data. Dep. Information Systems/Algoritmi R&D Centre University of Minho. Available at: http://www3.dsi.uminho.pt/pcortez/fires.pdf (accessed: 10.06.2019).

2. Biblioteki dlya glubokogo obucheniya: Keras. [Libraries for deep learning: Keras.] Open Data Science. Available at: https://habr.com /EN/company /ods/blog /325432/ (accessed: 10.06.2019) (in Russian).

3. Lutts, M. Izuchaem Python. [Learning Python.] St. Petersburg: Symbol-Plus, 2011, 1280 p. (in Russian).

4. Myuller, A., Gvido, S. Vvedenie v mashinnoe obuchenie s pomoshchyu Python. Rukovodstvo dlya spetsialistov po rabote s dannymi. [Introduction to machine learning using Python. A guide for professionals working with data.] St. Petersburg: OOO "Alfa-kniga", 2017, 480 p. (in Russian).

5. Gubenko, I.M., Rubinshteyn, K.G. Sravnitel’nyy analiz metodov rascheta indeksov pozharnoy opasnosti. [Comparative analysis of methods for calculating fire hazard indices.] Gidrometeorologicheskiy nauchno-issledovatel’skiy tsentr Rossiyskoy Federatsii. [Hydrometeorological Research Center of the Russian Federation.] Available at: http:// method.meteorf.ru /publ/tr/tr347 /gubenko.pdf (accessed: 10.06.2019) (in Russian).

6. Kin, B.A. Mashtabirovanie funktsiy s pomoshchyu scikit-learn. [Scaling functions using scikit-learn.] Available at: http:// benalexkeen.com/ feature-scaling-with-scikit-learn/ (accessed: 10.06.2019) (in Russian).

7. Esipov, Yu.V., Dzilyadzhi, M.S., Cheremisin, A.I. Logicheskoe i parametricheskoe modelirovanie predposylok i ustanovlenie mery opredelennosti realizatsii proisshestviya v siateme. [Logical and parametric modeling of perquisites and the establishment of measures of certainty of realization of the incident in the system.] Bezopasnost' v tekhnosfere 2017, vol. 6, no. 2, pp. 3-11 (in Russian).

8. Esipov, Yu.V., Mishenkina, Yu.S., Cheremisin, A.I. Modeli i pokazateli tekhnosfernoy bezopasnosti. [Models and indicators of technosphere safety.] Moscow: INFRA, 2018, 154 p. (in Russian).

Autors:

Filippenko Viktor Aleksandrovich,
student, Don State Technical University, (1, Gagarin sq., Rostov-on-Don, 344000, Russia),
philippenko@inbox.ru

Zotov Aleksey Vyacheslavovich,
student, Don State Technical University, (1, Gagarin sq., Rostov-on-Don, 344000, Russia),
azotzot@gmail.com