

Methodology for assessing water quality using a neural network

Evgeniy Schastlivtsev^{1,*}, *Natalia Yukina*² and *Ivan Kharlampenkov*²

¹Federal Research Center for Information and Computing Technologies, Kemerovo Branch in cooperation with the Institute of Water and Environmental Problems SB RAS, Russia

²Federal Research Center for Information and Computing Technologies, Kemerovo Branch, Russia

Abstract. A methodology for assessing the quality of surface and groundwater is proposed, using artificial intelligence methods, in particular, a fully connected neural network. Preliminary testing was carried out on water objects and a sufficiently high percentage of model accuracy was obtained.

1 Introduction

To assess the quality of surface and groundwater, a neural network was developed using the TensorFlow and Keras libraries [1 – 2], built into the Python programming language.

Supervised learning was used to create a neural network. As the correct answers, we used the results of calculations of hydrochemical data by the method of assessing water quality by associative indicators [3 – 6].

2 Materials and Methods

In accordance with the proposed method for choosing the correct answers, the water quality assessment is carried out according to two formulas (1) and (2). First, the normalized water indicators (NI) is determined (formula 1), and then the associative indicators of water composition (AI) (formula 2).

$$NI_i = \frac{C_i}{MPC_i} \quad (1)$$

Where, NI_i – normalized water indicators, C_i – arithmetic mean concentration of ingredient (i), mg/l; MPC_i – maximum permissible concentration of the corresponding ingredient (i), mg/l.

$$AI_{Nj} = \sum_{i=1}^N \frac{NI_i}{N} \quad (2)$$

* Corresponding author: schastlivtsev@ict.sbras.ru

Where, AI_{N_j} – associative indicators of water composition, NI_i – normalized water indicators arithmetic mean concentration of ingredient (i); N – number of ingredients.

3 Results and Discussion

Depending on the value of the AI indicator according to Table 1, the corresponding class of water quality is selected. Since, in the Kemerovo region, there are practically no waters corresponding to the first class of water quality, Table 1 was transformed by the author in order to optimize it for the use of a neural network, and contains only six classes of quality, in contrast to the previous methods.

Table 1. Water quality classes.

| Quality class | Name of classes | AI ranges |
|---------------|---------------------|-----------|
| 1 | clean | <1 |
| 2 | little contaminated | 1–2.5 |
| 3 | contaminated | 2.5–4 |
| 4 | dirty | 4–6 |
| 5 | very dirty | 6–10 |
| 6 | extremely dirty | >10 |

To improve the performance of the neural network, it is better to supply quality class numbers converted into vectors consisting of 1 and 0. For this, the encoding function - LabelBinarizer was imported from the Scikit-learn library [16]. As a result, the quality classes were converted to vectors:

- 1 - [1, 0, 0, 0, 0, 0],
- 2 - [0, 1, 0, 0, 0, 0],
- 3 - [0, 0, 1, 0, 0, 0],
- 4 - [0, 0, 0, 1, 0, 0],
- 5 - [0, 0, 0, 0, 1, 0],
- 6 - [0, 0, 0, 0, 0, 1]

Experimentally, the parameters of the neural network were selected (number of layers, number of neurons in each layer, activation functions, optimizer, number of Bath_Saze). As a result, we obtained the optimal architecture of the neural network: three fully connected (Dense) layers, on the first two layers there are 64 neurons and an activation function - relu, on the last - 6 neurons and an activation function - softmax (a screenshot of the program is shown in Fig. 1). As a result, a sufficiently high accuracy was obtained on the training (98.96 %) and test sample (96.63 %) (the screenshot is shown in Fig. 2).

Below is an algorithm for assessing water quality using the created neural network, tested on the example of hydrochemical data on the Chumysh River (Kemerovo Region) for 2011–2018:

- 1) loading a table with hydrochemical data for the Chumysh River;
- 2) normalization of hydrochemical data (subtract the mean and divide by the standard deviation);
- 3) launching a trained neural network;

- 4) obtaining the result of the neural network operation (determining the quality class of each sample and displaying the generalized result, in the form of the sum of the samples by class):
 - 1 class – 1 sample;
 - 2 classes – 17 samples;
 - 3 classes – 4 samples;

```

model = Sequential()
model.add(Dense(64, activation='relu', input_shape=(x_train.shape[1],)))
model.add(Dense(64, activation='relu'))
model.add(Dense(6, activation='softmax'))

print(model.summary())

```

Model: "sequential_4"

| Layer (type) | Output Shape | Param # |
|------------------|--------------|---------|
| dense_12 (Dense) | (None, 64) | 896 |
| dense_13 (Dense) | (None, 64) | 4160 |
| dense_14 (Dense) | (None, 6) | 390 |

Total params: 5,446
 Trainable params: 5,446
 Non-trainable params: 0

Fig. 1. Screenshot of neural network architecture for water quality assessment.

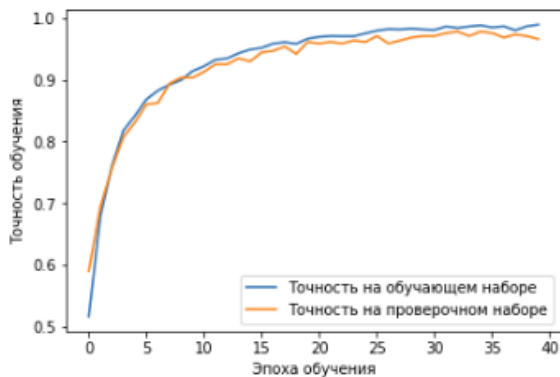


Fig. 2. Screenshot of graph of dependence of training accuracy on the number of training epochs.

- 5) choose the quality class, where most of the samples (in our case, it is the – 2 class);
- 6) determine the name of the water quality according to table 1, depending on the quality class.

As a result of data analysis, the water quality in the Chumysh river for 2011–2018 was determined, assessed as “little contaminated” (class 2).

According to the above algorithm, it is possible to assess the water quality of any water object.

4 Conclusion

To analyze the quality of surface and groundwater in the Kemerovo region, a methodology for assessing water quality using a fully connected neural network is proposed. Preliminary testing was carried out on water bodies and a sufficiently high percentage of accuracy was obtained, which indicates the adequacy of the proposed method. The results obtained by the neural network coincide with the results of traditional methods, such as the method for assessing the quality of water using the specific combinatorial index of water pollution (UCIPI) and can be used to assess the quality of techno-natural waters.

References

1. Tensor Flow library <https://www.tensorflow.org/>
2. Keras library <https://riptutorial.com/ru/keras/>
3. E.L. Schastlivtsev, N.I. Yukina, I.E. Kharlampenkov, Information-analytical system of geoecological monitoring of water resources of the coal-mining region, Bull. of KuzGTU, 2 (114):157–164 (2016)
4. E.L. Schastlivtsev, *Technogenic impact of coal mining enterprises on the environment (on the example of Kuzbass)*. Abstract dis. Doctor of Technical Sciences (Barnaul, 2006)
5. V.P. Potapov, V.P. Mazikin, E.L. Schastlivtsev, N.Yu. Vashlaeva, *Geoecology of coal-mining regions of Kuzbass* (Novosibirsk: Nauka, 2005)
6. V.A. Kovalev, V.P. Potapov, E.L. Schastlivtsev, Yu.I. Shokin, *Modeling of geoecological systems of coal mining regions* (KuzGTU. - Novosibirsk: Publishing house of the SB RAS, 2015)