GEOINFORMATICS 2021

EAGE

21117

The use of artificial neural networks to determine the type of rock saturation

***A. Stolpakov** *(Taras Shevchenko National University of Kyiv)*

SUMMARY

In this stated theses about application of artificial neural networks for a problem of classification of rocks on saturation on such classes, as brine or oil saturation on the basis of parameters received during researches by a method of nuclear magnetic resonance. Based on them, the following results were obtained, the accuracy of correct answers is 93% when adjusting the resonance of the echo spacing 0.25 m/s. The following conclusions were drawn about the need to exclude nuclear magnetic resonance from research if these data will be used for this task.

Introduction

Relevance: Because laboratories conduct a lot of research, including the method of nuclear magnetic resonance. They have accumulated a lot of data without conclusions. Therefore, the problem of finding different patterns of these data for their further use in solving various problems is relevant. In particular, modern methods, which now do not take as much time as it took before.

Objective: Based on the input parameters of nuclear magnetic resonance studies with a high probability to correctly determine what exactly is a saturated rock with brine or oil.

Overview of existing views on the problem: In general, of course, it is now very popular to use neural networks wherever possible. But few studies use this to determine the type of saturation based on nuclear magnetic resonance. These studies are generally used to assess porosity and similar problems. (Jorand et al., 2011; Mukhametdinova et al., 2020)

Problem statement: Choose the type of artificial neural network that will give the best results in the problem of determining data patterns of behavior, to classify data into 2 classes by rock saturation, in our array there are only two types of saturation: brine and oil.

Theory

After analyzing the main types of neural networks used, I chose a direct-flow neural network, (Ghareb et al., 2018) which is well implemented in the software product Matlab R2019a from company Mathworks. In order to understand what a direct-flow artificial neural network, first of all is understanding what these neural networks are, and then understand exactly how this species works.

(Walczak and Cerpa, 2003) «Artificial neural networks are a technology based on studies of the brain and nervous system. These networks emulate a biological neural network but they use a reduced set of concepts from biological neural systems. Typically, the neutrons are arranged in a layer or vector, with the output of one layer serving as the input to the next layer and possibly other layers. A neutron may be connected to all or a subset of the neutrons in the subsequent layer, with these connections simulating the synaptic connections of the brain. Weighted data signals entering a neutrons simulate the electrical excitation of a nerve cell and consequently the transference of information within the network or brain. The input values to a processing element, i_n, are multiplied by a connection weight, $w_{n,m}$, that simulates the strengthening of neural pathways in the brain. It is through the adjustment of the connection strengths or weights that learning is emulated in ANNs. All of the weight-adjusted input values to a processing element are then aggregated using a vector to scalar function such as summation ($y = \Sigma w_{ij} x_i$), averaging, input maximum, or mode value to produce a single input value to the neutrons. The transfer function transforms the neutrons input value. Typical this transformation involves the use of a sigmoid, hyperbolic-tangent, or other nonlinear function. The process is repeated between layers of processing elements until a final output value, o_n, or vector of values is produced by the neural network».

It has only 3 input layers has as many neurons as the input data. One hidden layer. But the last layer is the source, it already gives the final answer as to whether the input belongs to a certain class. That is, the number of neurons in this layer is exactly as many as we can have classes. In addition to the different types of neural networks, which differ in the placement of neurons in different forms, and how they interact with the data. There are also different types of training of these same networks, setting the best weights mentioned above. There are three types of learning supervised, unsupervised and semi. The first type is when the training dataset is marked, respectively unsupervised is not marked and semi where part is marked and part is not. (Ben Dickson, 2021) This paper uses supervised machine learning.

In Figure 1 we see the window of the result of the network in Matlab, where you can see how many epochs have passed, how many neurons at the hidden level. You can also build a performance, training state, error histogram, confusion, receiver operating characteristic.

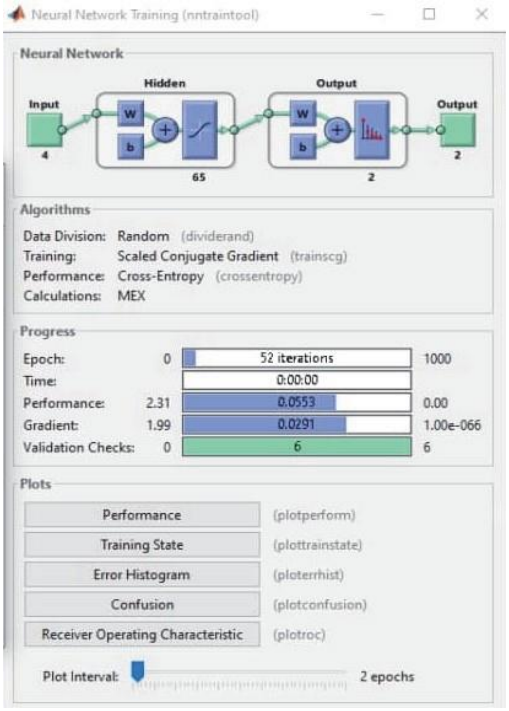

Figure 1 *Image piece of window in matlab with the best results. We have 65 neutrons in hide layer*

Now it is necessary to understand on the basis of what we will do this analysis that is simply to understand what is NMR. (Hemminga, 1992) «Nuclear magnetic resonance (NMR) is a spectroscopic technique based on the magnetic properties of atomic nuclei. The principle of NMR was predicted in the mid1930s by the Dutch physicist (Gorter et al., 1936), and, using a beam-splitting technique, (Rabi et al., 1938) performed the first resonant experiment on silver atoms. However, it was not until 1945 that the American physicists (Bloch et al., 1946) and (Purcell et al., 1946) actually discovered NMR in the form in which it is known today. For this work they received the Nobel prize for physics in 1952. Initially, NMR was the domain of physicists, but when in 1950 by Proctor and Yu discovered the chemical shift, it became clear that NMR could play an important role in chemistry and other areas of science. As NMR spectroscopy reveals useful information about molecular structure, chemical reaction rates and diffusion processes, it has rapidly progressed to become the most powerful non-destructive analytical tool in chemistry. In biochemistry, NMR spectroscopy has made contributions to elucidating the structures of cellular membranes, nucleic acids, proteins and viruses».

The essence of the work

At the very entrance, received the input data in the form of pictures, they were distributed by wells. They have been digitized to Excel spreadsheets also by wells. Then regrouped by parameters in separate database tables. Then again regrouped into three new tables where they were distributed by the parameter of echo spacing, because data have three different of this 0.25m/s, 1 m/s and 5 m/s research parameter nuclear magnetic resonance. The final tables have a structure where each column is an NMR parameter and the row is well data. But preprocessing was done before the merger. Distribution charts have been constructed to assess the data itself, how polluted they are with emissions, and to remove these emissions. Lines where at least one parameter was absent were also removed. Figure 2 is good example of how manual chart analysis improves data distribution: For the FFI brine saturation, 5m/s of echo spacing, 2 values in the diagram tail were removed. As a result of cleaning, the data on two wells were deleted, and the diagram itself was improved many times.

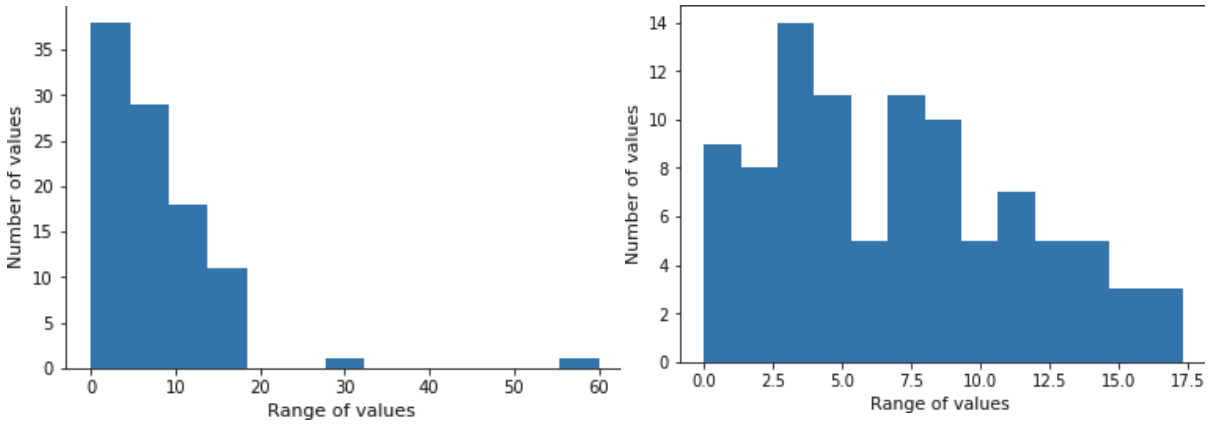

Figure 2 *The difference between the distribution diagrams before cleaning on the left and after on the right.*

After clearing and merging the data, they were converted to csv format. Before submitting them to the network itself, they are normalized to a minimax normalization. A little about this type of normalization: (Han et al., 2012). «Min-max normalization performs a linear transformation on the original data. Suppose that min (d) and max (d) are the minimum and maximum values of an attribute, D. Min-max normalization maps a value, of an attribute, D. Min-max normalization maps a value, data in the range [new min (d), new max(d)] by computing

$d = \frac{d - \min(d)}{\max(d) - \min(d)}$ Where d is all elements in data. This element-by-element operation

Minimax normalization preserves the relationships among the original data values. It will encounter an "out-of-bounds" error if a future input case for normalization falls outside of the original data range for A.»

After normalization of the data it is necessary to give still values of classes as supervised machine learning is applied here. The last column contains information about the membership of a particular row. In the form of 0 and 1. Where 1 is the saturation of oil and 0, respectively, brine saturation. But due to the peculiarities of matlab, these data need to be modified. The matlab at the input for class values takes as many columns as there should be classes. Accordingly, this provides good opportunities when you do not have a fully expressed class, you can specify for example 0.9 and 0.1 in one column. That is 90% of belonging to 1 class and 0.1 to second. But not in this case the classes are completely defined here. Therefore, you just need to fill the second column with opposite values. This was done as follows: to create a new value in the corresponding line, take the value of the same line and subtract one. Next, the main essence of the work was to find the best parameters of the neural network to obtain the best results of input data processing. That is to find such combination of parameters at which we will receive the greatest percent of correct answers, and in our case it is correct distribution on classes on saturation of rocks.

In the Matlab software product, we can change parameters such as the number of input neurons, not the number of neurons, and manually select the parameters we want, find patterns to solveour problem. In my case, the best input was 4 input neurons, each of which was responsible for a separate parameter of one sample. Then in the program window you can specify the number of neuronsat the hidden level. The number of neurons in the range from 60 to 85 proved to be the best. Then thereis an initial layer of a neural network their quantity makes as many as it is necessary to define classes. We can also set a percentage for the validation and test sample of the total input. The validation samplemakes sure that the neural network is not retrained, when the maximum accuracy of 100% is not achieved, the training continues, but when the validation sample makes 6 correct answers, the trainingis stopped because it is considered that at this point it has reached its best result. And the test sample is already after the cessation of training testing of the neural network for accuracy. Not all of these samples are related. But it was not enough to find good parameters. A bit of luck will not hurt because it is a stochastic process (means random), and so it is because we choose only the percentage of data on which

the neural network will be tested, validated (it means avoid retraining) and tested, and the order of data submission to the network we do not we ask that they are given randomly so that the neural network does not adapt to some data, and on others it will not give a good result.

Conclusions

My conclusions can be divided into two sections, the first section, on the parameters that were established as the best for finding patterns in data for the distribution of the saturation of rocks with brine or oil. So, the optimal number of neurons on the hidden layer is from 65 to 85. The input number of parameters is 4 because it reduces the accuracy less, because a smaller number of them reduces the accuracy, and increasing their number does not increase, but rather reduces the percentage of class determination. The obtained accuracy is 93% of the correct answers.

Another section on the application of the results. At the input I had 8 parameters of nuclear magnetic resonance with 3 different echo spacing settings 0.25, 1 and 5m/s. The number of good parameters that can be used to train the neural network, as mentioned above is 4. That is, the remaining parameters to determine the type of rock saturation can be omitted. Which makes it cheaper to perform these works. Also the percentage of correct answers of 93% was obtained by setting 0.25 m/s of echo space also a good percentage of 90 at 5 m/s echo of space, and at 1 m/s the best figure that was achieved is 77%, which is not good. Thus, research at 1 m/s echo space to solve the problem of rock saturation can also not be done, because the exact is not very good, which in turn reduces the cost of the process, and the result does not worsen. That is, it is quite possible to apply this network not only on the core in the laboratory, but in the well for already besieged wells, or new ones.

References

Ben Dickson [2021] What is semi-supervised machine learning? 1p.

Bloch, F. [1946] Nuclear Induction. Physical Review, 70(7-8), 460–474p.

Gorter, C.J. [1936] Physica 3, Book. 995-998 p.

Ghareb, H., Elsakka Ahmed, Nyein Chaw [2018] Artificial Neural Network (ANN) Prediction of Porosity and Water Saturation of Shaly Sandstone Reservoirs, Article. 1-3p.

Han, J., Kamber, M., and Pei, J. [2012] Data Preprocessing. Data Mining, 83–124. Article, 31-33 p.

Hemminga, M.A. [1992] Introduction to NMR. Trends in Food Science & Technology, 3, 179–186 p.

Jorand, R., Fehr, A., Koch, A. and Clauser C. [2011] Study of the variation of thermal conductivity with water saturation using nuclear magnetic resonance, J. Geophys. 1-3 p.

Mukhametdinova, A., Mikhailova, P., Kozlova, E., Karamov, T., Baluev, A. and Cheremisin, A. [2020] Effect of Thermal Exposure on Oil Shale Saturation and Reservoir Properties 1-3p.

Proctor, W.G. and Yu, F.C. [1950] The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound. Physical Review, 77(5), 717–717p.

Purcell, E.M., Torrey, H.C. and Pound, R.V. [1946] Resonance Absorption by Nuclear Magnetic Moments in a Solid. Physical Review, 69(1-2), 37–38p.

Rabi, I.I., Zacharias, J.R., Millman, S. and Kusch, P. [1938] A New Method of Measuring Nuclear Magnetic Moment. Physical Review, 53(4), 318–318 p.

Walczak, S. and Narciso Cerpa [2003] Encyclopedia of Physical Science and Technology (Third Edition), I.A Biological Basis of Artificial Neural Networks. Book 631-645p.