Classification method of carbonate reservoir welltest interpretation model based on long and short-term memory network

Jinghan Lu, Minglu Wu *
China University of Petroleum, QingDao, China

*Corresponding author e-mail: z18020002@s.upc.edu.cn

Abstract. The development of carbonate reservoirs is inseparable from the measurement of formation parameters, and well test interpretation techniques are often used to determine these parameters. Determining the carbonate reservoir type according to the shape of the well test interpretation curve will help people better understand the reservoir situation and enable more precise well test interpretation. This research focusing on the well test interpretation of carbonate oil and gas reservoirs combined with the actual data of the mine field and uses deep learning algorithms such as feature value extraction, clustering and neural network training to classify the types of carbonate oil and gas reservoirs. Neural network training is further carried out for different types of carbonate reservoirs, so as to achieve the purpose of identifying reservoir types at the algorithm level.

1. Introduction
Carbonate oil and gas reservoirs occupied a pivotal position in the world's oil and gas reserves, and the development of carbonate reservoirs is inseparable from the determination of stratigraphic parameters. The understanding of carbonate reservoirs and the production performance of oil wells Based on the analysis of change characteristics, it is of great significance to effectively use the existing well test data to obtain reservoir information to the greatest extent and guide the development of this type of oilfield. In the well test interpretation method, the judgment of the reservoir type is a crucial step. At present, this step is mostly manually judged by experienced practitioners, which consumes a lot of manpower and is relatively inefficient.

Therefore an operation processes that can be separated from human operations and easily distinguish the types of carbonate reservoirs is particularly important.

2. Long Short-term Memory Networks
In the field of machine learning, when we need to process some data related to time series, we usually use the recurrent neural network (RNN) model [1], which is usually used to process the natural language learning (NLP) field and related question. In the process of oil and gas field development, the well test curve records the data of pressure and production changes over time. Strictly speaking, each data point is related to all previous data points (excluding data points that contain noise). It also determines all subsequent data points. The connection between nodes in this time series happens to be the problem that the Recurrent Neural Network (RNN) model is best at dealing with [2].
The long short term memory network, long short term memory abbreviated as LSTM [3], is specifically designed to solve the problem of the traditional recurrent neural network (RNN) in dealing with the long dependency problem [4] that the model is too complex. In order to facilitate the derivation of the formula, we introduce the schematic diagram of the LSTM neural network unit shown in Figure 1. Use h to represent the output of the LSTM unit, x to represent the input, and c to represent the value of the memory unit. Then the entire LSTM unit can be updated according to the following steps.

![Schematic diagram of LSTM neural network unit.](image)

The first step is to calculate the value of the candidate memory unit $\tilde{c}_t$ at the current moment according to the above-mentioned RNN forward propagation calculation formula. $W_{xc}$ and $W_{hc}$ in formula represent the input data of the previous layer and the output of the LSTM unit at the previous time $t$. Weight:

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

Because in LSTM model, in addition to the current output data $x_t$ and the output value $h_{t-1}$ of the LSTM unit at the previous time, the calculation of all "gates" must also consider the impact of the memory unit $C_{t-1}$ value at the previous time. This is also the source of the "memory" feature in the LSTM network model, so there is an input gate calculation formula:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$

Similarly, the forget gate $f_t$ is used to limit the influence of the memory information on the current unit value, and to prevent the network from remembering too much redundant information to affect the training effect:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

After integrating the calculation of the input gate and the forgetting gate, add a nonlinear activation function $\sigma$ to the output (usually using the Sigmod function or the ReLU function), and the output of the output gate can be obtained:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o)$$

The structural design of the three gates and the independent Cell unit enable the LSTM neural network model to have the ability to save, read and update historical data on a longer-term sequence scale.

3. Model training and discussion
We select the carbonate reservoir well test curve of block M as the data set. Then we use kmeans algorithm to classify the data set and add tags and use GAN algorithm to expand the data set.
Based on the Keras package, a LSTM long and short-term memory neural network model is built to identify the type of carbonate reservoirs. The LSTM model built in this example contains an LSTM unit layer and an output layer. The output layer corresponds to the eight well test curves divided in this article. Category, select the Softmax function as the activation function.

The model is iterated 500 times in total, and the final loss value on the test set is 1.3294 and the hit rate is 0.9485.

**Figure 2.** The learning curve of the LSTM long and short time neural network model on the training set.

**Figure 3.** The learning curve of the LSTM long- and short-term neural network model on the test set.
Figure 2 and Figure 3 are the learning curves on the training set and the test set when using the carbonate reservoir test curve data set LSTM long-term neural network model. The number of iterations is 500, and the red curve represents the model. The drop curve of the loss function, the green curve is the hit rate of the model on the corresponding data set. It needs to be pointed out that due to the different loss functions of different models, the value of the loss function does not have practical significance. The change trend of the loss function is of guiding significance for model training, especially the loss function in the fully connected deep neural network. Differentiate.

It can be seen from the figure that whether it is the test set or the training set, after about 100 iterations, the change range of the curve begins to slow down, and gradually converges at about 150 iterations. The final accuracy rate on the test set is about 95%. In particular, after the model converges, there are still three clusters of relatively concentrated peaks appearing on the curve, which may be caused by the local overfitting of the model. Since LSTM is essentially a cyclic neural network, this deviation will be affected in a certain period. Keep repeating, but it can be seen that the time interval of this repetition itself is constantly getting larger, that is, the forgetting mechanism of the LSTM network is working.

Figure 4. Changes in the accuracy of the LSTM long- and short-term neural network model on the training set and test set

In Figure 4, we moderately intercepted the part of the accuracy curve greater than 50%. When zooming in on the entire curve, we can find that the LSTM network model has approximately met a local optimal solution when iterated to 50 iterations. The accuracy rate increased sharply around this point, and then gradually flattened, and after 200 iterations Complete convergence.

4. Conclusion
For carbonate reservoirs with more complex structures, by drawing well test curves, applying kmeans algorithm for clustering and using GAN to expand the data set, successfully realized the type identification of well test curves by building the LSTM model, and on the test set the effective rate is 94.85%. The algorithm can quickly converge to a relatively stable value, but it will fluctuate periodically. It is speculated that using data with a larger time span can make the model converge in a more stable space.
References

[1] Williams R, Zipser D. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks [J]. Neural Computation, 2014, 1(2):270-280.

[2] Yang Li, Wu Yuqian, Wang Junli, et al. Overview of cyclic neural network research [J]. Computer Applications, 2018, 38(S2): 1-6+26.

[3] Hochreiter S, Schmidhuber J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[4] Wang Xin, Wu Ji, Liu Chao, et al. Fault time series prediction based on LSTM recurrent neural network [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(04): 772-784.