PAPER • OPEN ACCESS

Prediction on Production of Oil Well with Attention-CNN-LSTM

View the article online for updates and enhancements.

Prediction on Production of Oil Well with Attention-CNN-LSTM

Shaowei Pan[1*], Jizhe Wang[1] and Wei Zhou[2]

[1] School of Computer Science, Xi'an Shiyou University, Xi'an, 710065, China

[2] Shengli Oilfield Exploration and Development Research Institute, SINOPEC, Dongying, 257100, China

*Corresponding author's e-mail: swpan@xsyu.edu.cn

Abstract. The production prediction of an oil well is of great significance for realizing the intelligent monitoring of oil wells and improving oil recovery. In order to overcome the shortcomings of existing methods in the production prediction of oil wells, this study proposes a hybrid deep learning model based on the attention mechanism by combining the convolutional neural networks and the long short-term memory neural networks (Attention-CNN-LSTM). First, Attention-CNN-LSTM extracts the fine-grained features of the input data through CNN. Second, it obtains the coarse-grained features through LSTM and realizes the optimization of the results by the different features of the input data through attention mechanism. The Attention-CNN-LSTM model is trained and tested by using the production datasets of T1 well and T2 well of an oilfield in the southern China, respectively, and the prediction results of Attention-CNN-LSTM on the test set are compared with the prediction results of back propagation neural networks (BP), support vector regression (SVR), LSTM, Attention-LSTM and CNN-LSTM on the test set. The final comparison results show that Attention-CNN-LSTM has the best prediction performance and can achieve more accurate prediction of oil production.

1. Introduction

As everybody knows, oil is a very important source of energy. Therefore, the production prediction of an oil well has always been a hot topic in the petroleum industry. Oil is produced by oil wells. In addition to the geological factors, the production of an oil well is also affected by many external factors such as time, geographic location and production equipment. So it is difficult to accurately predict the production of an oil well by using the curve analysis method[1] and the mathematical modeling method[2]. However, it is very necessary to achieve accurate and efficient production prediction for an oil well. Because an accurate and efficient production prediction model of an oil well can greatly save the manpower and material resources consumed in the oil extraction process, reduce the unnecessary expenditures, and improve the economic benefits for oilfields.

The current data-driven methods for production prediction of an oil well mainly include the traditional machine learning methods and the deep learning methods. The traditional machine learning methods include autoregressive integrated moving average model (ARIMA), SVR, intelligent optimization algorithm and BP neural networks[3-6]. They have simple basic principles and are widely used, but they all show certain deficiencies in the production prediction of an oil well: ARIMA has the high stability requirement on production data of an oil well; SVM have the origin shortcomings in processing large data samples; The intelligent optimization algorithm has the higher

requirement for parameter settings; BP can easily reach the local minimum, such that some cannot obtain the optimal solution, and generalization ability is poor.

CNN and RNN play an important role in deep learning nowadays. CNN has been successfully applied in image recognition in the past[7-8]. RNN has excellent performance in natural language processing. But when processing the long-term sequence data, RNN always shows gradient vanish or gradient explode[9]. In order to solve the above problems, Hochreiter and Schmidhuber proposed LSTM[10]. Since LSTM can solve the problem of long-term dependence, it has obtained certain applications in the fields of production prediction for an oil well [11-13]. However, if LSTM is used solely, it is not stable enough in long-term sequence data processing. The importance of time series characteristics in most time series data is different. And ignoring some important time series characteristics may result in a decrease in the prediction accuracy of a prediction model. The attention mechanism can solve the above problem. It can make the prediction model focus on the most effective characteristics of time series data with limited resources[14-16]. Therefore, we propose the Attention-CNN-LSTM model for the production prediction of an oil well in this paper. With Attention-CNN-LSTM, the different characteristics that affect the oil well production can be noted and processed, and the noises in the production data of an oil well can also be removed. The oilfield researchers can accurately monitor the production of an oil well by using Attention-CNN-LSTM, and then provide an important technical reference for the formulation of the next production plan of oilfields.

Our contributions are as follows:
- We introduced an Attention-CNN-LSTM model to overcome the challenges of the production prediction for an oil well, and established a model which is both efficient and accurate.
- The production prediction of an oil well with Attention-CNN-LSTM was performed and compared with five other existing methods including BP, SVR, LSTM, Attention-LSTM and CNN-LSTM with respect to the collected actual datasets. We show that our proposed method could obtain state-of-the-art results in comparison with these methods.
- We gave the reasons why Attention-CNN-LSTM performs best in the production prediction of an oil well compared to other machine learning methods.

2. Related works

In the early stage of oilfield development, the curve analysis method and the mathematical modeling method were widely used in the production prediction of an oil well [1-2]. Both the curve analysis method and the mathematical modeling method establish a mathematical model for the production prediction of an oil well through nonlinear regression, and then make predictions for the production of an oil well in the future. The applications of the curve analysis method and the mathematical modeling method are very simple, but their prediction accuracy is low. When the traditional machine learning methods are used to predict the production of an oil well, the training samples are used to train the machine learning models [3-6]. After the training is completed, the machine learning models are tested, and they can be used for prediction if the test is qualified. But the traditional machine learning methods generally require that all data should be put into the memory during training, so when the amount of training data increases, their performance will be greatly reduced. Deep learning is a method of machine learning based on the data characterization and learning, its motivation is to establish and simulate a neural network for analysis and learning of the human brain. CNN, RNN and the generative adversarial networks (GAN) are three typical deep learning algorithms, they have the advantages of strong learning ability, wide coverage, good adaptability and portability. At present, LSTM is used in the production prediction of an oil well and has achieved good results [11-13]. However, due to the harsh underground production environment, the oil production data of oil wells often contains multiple noise components, which are non-linear and non-stationary time series data. Obviously, there must be deficiencies when using the single LSTM to predict the production of an oil well. Therefore, this paper combines CNN, LSTM and the attention mechanism to construct a production prediction model for an oil well.

3. Methodology

3.1. Problem formulation

The objective of production prediction for an oil well is to predict the current and future changes in the production of an oil well according to the multivariate factors from the past. Formally, the input is represented as the fully observed feature vector set $\{x_t\}_{t=1}^{T} = x_1, x_2, \cdots, x_T$ and the corresponding production of an oil well $\{y_t\}_{t=1}^{T} = \{y_1, y_2, \cdots, y_T\}$, where T is the length of total time steps, such as the number of days in the collected dataset. At time step t, y_t is the oil well production, and x_t is typically the vector of multivariate factors.

The production prediction problem of an oil well uses the time series of multivariate factors $\{x_t\}_{t=1}^{T}$ and the actual oil well production $\{y_t\}_{t=1}^{T}$ as inputs and constructs a model F to predict y at future time steps:

$$\{\hat{y}_t\}_{t=T+1}^{T+\Delta} = F(\{x_t\}_{t=1}^{T}, \{y_t\}_{t=1}^{T}) \tag{1}$$

This formulation is different from that in autoregressive models, which usually assume that $\{x_t\}_{t=T+1}^{T+\Delta}$ is available when predicting $\{\hat{y}_t\}_{t=T+1}^{T+\Delta}$, because they are designed to model mapping between the conditions and consequences.

3.2. CNN

Different from the traditional fully connected neural networks, the local connection and weight sharing are the two major characteristics of CNN [7-8]. CNN is generally composed of the input layer, the convolutional layer, the pooling layer, the full connection layer and the output layer. The core of CNN is the convolutional operation. With the advantages of convolutional operation, CNN can abstract and express the original oil production data at a higher level. In this paper, the features of the original oil production data are processed by CNN, the correlation between the multi-dimensional oil production data is mined and the noises are removed, and the processed data information with high-dimensional characteristics is introduced into LSTM as a whole to predict the time series of oil well production.

3.3. LSTM

Compared with RNN, the key to LSTM is the introduction of self-circulation, which generates a continuous flow path for gradient for a long time [10]. Another key extension of LSTM is to make the weights of self-circulation context-dependent, rather than fixed. Therefore, LSTM overcomes the problems of gradient disappearance and gradient explosion that RNN has.

3.4. Attention mechansim

The attention mechanism can be used to extract the salient features in the sub-sequences of long-time sequence and applied to calculate the weighted summation for the vector expression of hidden layer of LSTM output, which realizes the efficient allocation of information resources. The specific formulas are as follows.

$$c_i = \sum_{i=1}^{k} a_i h_i \tag{2}$$

$$s_i = \tanh(w^T h_i + b_i) \tag{3}$$

$$a_i = soft\max(s_i) \tag{4}$$

In the above formula (2), a_i is the accumulated attention weight of state variables, h_i is the input feature vector, and c_i is the environment vector obtained by the weighted summation of the input feature vector and the attention weight. In the above formula (3), w^T is the output hidden layer vector of LSTM, b_i is the offset term, and s_i is the score of each hidden layer vector in LSTM. In the above formula (4), function *SoftMax* is used to normalize s_i to get the final weight coefficient.

3.5. Attention-CNN-LSTM

The basic structure of Attention-CNN-LSTM proposed in this paper is shown in Figure. 1. As can be seen from Figure. 1, Attention-CNN-LSTM is composed of the input layer, the CNN module, the LSTM module, the attention mechanism module and the output layer. The basic execution process of Attention-CNN-LSTM is as follows:

- The input data are received by the CNN module, and then the spatial features of the multivariate time variables are extracted by the convolutional layer and the pooling layer, which are transmitted to the LSTM module after the noises are removed;
- The irregular time information model is established by using the spatial characteristics of the input multivariate time variables in the LSTM module;
- The weight allocation is fitted automatically through LSTM in the attention mechanism module, and the important feature components are given greater weight;
- The hidden layer output vectors of different time nodes in the LSTM module and the corresponding weights are multiplied and summed, which is used as the final weight of Attention-CNN-LSTM;
- The fitting of oil well production is completed in a fully connected network structure with Attention-CNN-LSTM.
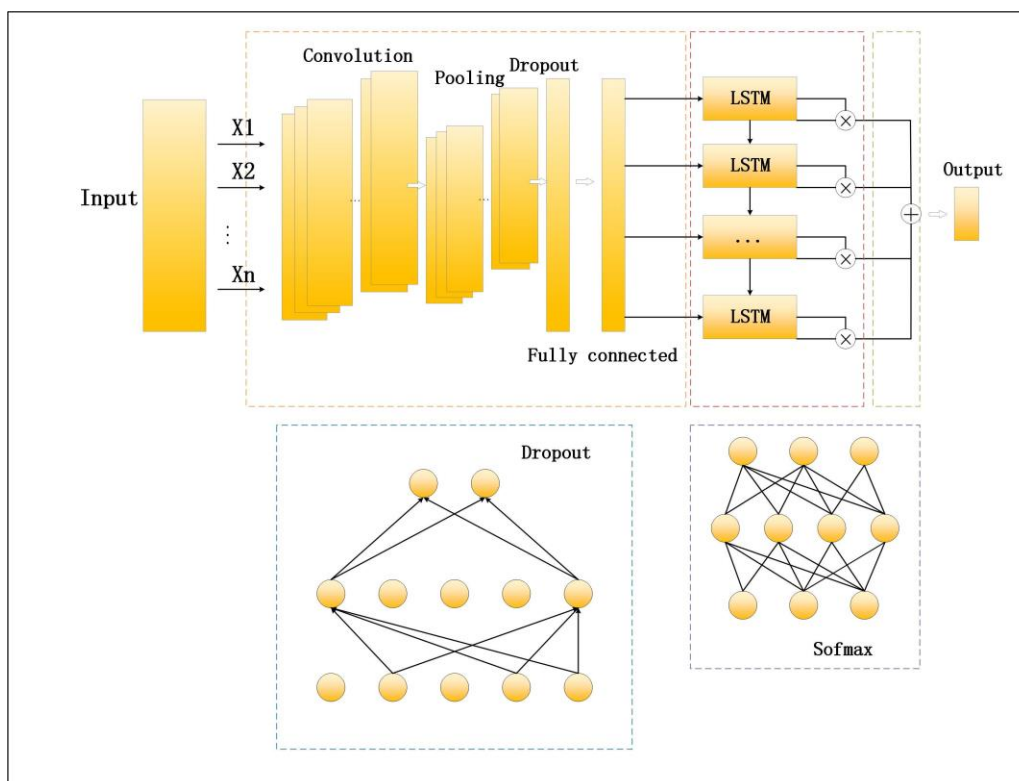

Figure 1. Structure diagram of Attention-CNN-LSTM.

4. Experiment and discussion

4.1. Data Description

The oil wells studied in this paper are located in an oilfield in southern China and include the T1 well and T2 well. For this oilfield in southern China, these two oil wells are the main producers of crude oil. The T1 well was put into development in January 1996, and the T2 well was put into development in July 1996. This paper collected production data of the T1 well and T2 well (see Table 1 and Table 2). In Table 1 and Table 2, the time step is set to a day to build the evaluation dataset for these two oil wells. According to the specific development characteristics of these two oil wells and considering the data integrity for the deep learning model, four parameters with great impact on the production of these two oil wells were determined. As shown in Table 1, there are six columns of data in this dataset. The first column is the date, and the second column is the oil production of the T1 well for the corresponding date. The third column is the effective thickness of the oil layer in the T1 well, the fourth column is the diameter of the pump in the T1 well, the fifth column is the pump efficiency, and the sixth column is the formation coefficient, which is closely related to crude oil production in the T1 well for the corresponding date. The meanings of data in Table 2 is the same as in Table 1, so they are not reinterpreted.

Table 1. Daily oil production data of T1 well.

Time period	Oil well production (t)	effective thickness of the oil layer (m)	pump diameter (mm)	pump efficiency (%)	formation coefficient (mD.m)
1996-01-01	316.6	22.8	38	64.0	0.85
1996-01-02	316.9	22.8	38	63.9	0.85
1996-01-03	317.1	22.8	38	63.8	0.85
…	…	…	…	…	…
2017-12-29	36.3	15.8	38	56.4	1.00
2017-12-30	36.5	15.8	38	56.2	1.00
2017-12-31	36.7	15.8	38	56.0	1.00

Table 2. Daily oil production data of T2 well.

Time period	Oil well production (t)	effective thickness of the oil layer (m)	pump diameter (mm)	pump efficiency (%)	formation coefficient (mD.m)
1996-07-01	222.3	9.8	38	28.0	0.85
1996-07-02	223.3	9.8	38	27.9	0.85
1996-07-03	224.3	9.8	38	27.8	0.85
…	…	…	…	…	…
2013-12-29	56.9	8.7	38	87.3	1.00
2013-12-30	56.6	8.7	38	87.1	1.00
2013-12-31	56.4	8.7	38	87.0	1.00

4.2. Measure Metric

In this study, three measures of prediction accuracy are calculated using the acquired predicted values, the RMSE, MAE, and MAPE, which are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{7}$$

The model with the lowest values in these measures is considered the best prediction model.

4.3. Implementation details

After many attempts, the main parameters of Attention-CNN-LSTM are determined as follows (Table 3).

The input of Attention-CNN-LSTM is set to the production and related influencing factors of an oil well at time $t-1(t=1,2,\cdots,12)$, and the output is set to the oil well production at time t. Then the datasets of T1 well and T2 well are divided into the training dataset and the test dataset by

taking the day as time step, respectively. The training dataset is used to train the Attention-CNN-LSTM model, and the test dataset is used to test the Attention-CNN-LSTM model after training. Finally, the test results are represented by the three indexes, RMSE, MAE and MAPE. In addition, the same training dataset and test dataset are used to train and test the five machine learning methods of BP, SVR, LSTM, Attention-LSTM and CNN-LSTM, respectively, and the test results are also characterized by RMSE, MAE and MAPE. At the same time, in order to avoid the error caused by the weight initialization of neural networks, BP, SVR, LSTM, Attention-LSTM, CNN-LSTM and Attention-CNN-LSTM are trained and tested for 30 times based on the same input and output. Thirty different groups of RMSE, MAE and MAPE are obtained, and their average values are taken as the final experimental results, as shown in Table 4.

Table 3. Settings of the main parameters of Attention-CNN-LSTM.

Parameter	Value
Epoach	50
Batch size	30
CNN filter Size	64
CNN Kernel_size	3
CNN activation	Relu
CNN pool_size	2
LSTM Unit Number	64
LSTM droupout	0.2
Dense	1
Loss	Mae
Optimization Function	Adam
Learning rate	0.1
Patience	4

Table 4. RMSE MAE and MAPE comparison of the T1 well and T2 well.

Oil well	Method	RMSE	MAE	MAPE
T1	BP	2.167	1.579	3.532
	SVR	0.786	0.850	1.783
	LSTM	0.927	0.651	2.131
	Attention-LSTM	0.728	0.606	0.020
	CNN-LSTM	1.310	1.105	0.039
	Attention-CNN-LSTM	0.315	0.246	0.008
T2	BP	1.875	1.218	2.113
	SVR	0.590	0.912	0.778
	LSTM	0.653	0.492	0.802
	Attention-LSTM	0.583	0.441	0.008
	CNN-LSTM	0.565	0.435	0.008
	Attention-CNN-LSTM	0.402	0.303	0.005

4.4. Results and discussion

The prediction results of BP, SVR, LSTM, Attention-LSTM, CNN-LSTM and Attention-CNN-LSTM on the test dataset of the T1 well and T2 well are shown in Table 4, respectively. As can be seen from Table 4, for T1 well, the RMSE, MAE and MAPE generated by Attention-CNN-LSTM are the smallest, which are 0.315, 0.246 and 0.008, respectively; the largest RMSE, MAE and MAPE are generated by BP, which are 2.167, 1.579 and 3.532, respectively. For T2 well, the RMSE, MAE and MAPE generated by Attention-CNN-LSTM are also the smallest, which are 0.402, 0.303 and 0.005, respectively; the largest RMSE, MAE and MAPE are also generated by BP, which are 1.875, 1.218 and 2.113, respectively. It can be found from Table 4 that the prediction performance of the combined deep learning models is better than that of the traditional machine learning models and single deep learning model. Among the three combined deep learning models, the prediction performance of Attention-CNN-LSTM is the best. This is because compared with other deep learning models, the CNN module in Attention-CNN-LSTM reduces the amount of data calculation and retains the significant information of data, while the LSTM module solves the problems of gradient disappearance or gradient explosion that may occur, and the attention mechanism module highlights the effective features related to prediction by giving more weight to the important feature components.

5. Conclusions

(1) Compared with BP, SVR, LSTM, Attention-LSTM and CNN-LSTM, Attention-CNN-LSTM is more suitable for predicting the time series data such as oil well production.

(2) Attention-CNN-LSTM can not only use the attention mechanism to obtain the important features that affect an oil well production, but also extract the high-dimensional features of input data through CNN, and can also predict the time series of input high-dimensional features with LSTM, avoiding the phenomenon of gradient disappearance or gradient explosion that occurs when RNN processes the long-term sequences.

(3) This paper creatively introduced Attention-CNN-LSTM into an oil well production with the hopes that this idea could serve as inspiration or as a technical reference for future deep learning and even machine learning methods in the petroleum industry.

References

[1] Q. Yu. (1992) A technique for constructing oil well index curves in a naturally fractured reservoir. Petroleum Exploration and Development. Beijing, vol. 19, pp. 69–78.

[2] S. Yu, Y. Wang, J. Wang, and X. Li. (2001) A study on economic limit production from wells by bailing. Petroleum Exploration and Development. Beijing, vol. 28, pp. 96–97.

[3] J. Gu, G. Sui, Z. Li, W. Liu, Y. Wang, Y. Zhang, and W. Cui. (2018) Oil well production forecasting method based on ARIMA-Kalman filter data mining model. Journal of Shenzhen University (Science and Engineering). Shenzhen, vol. 35, pp. 575–581.

[4] Y. Yang, Y. Cheng, Y. Qiao, and J. Liu. (2018) A study on oil well production prediction based on Time Series dynamic analysis. Journal of Southwest Petroleum University (Science & Technology Edition). Chengdu, vol. 42, pp. 82–88.

[5] X. Xu, H. Zhou and S. Xu. (2016) Support vector machine model and beam pumping unit pumping parameters optimization. Oil Field Equipment. Shanghai, vol. 45, pp. 1–5.

[6] C. Li, M. Tan, and K. Zhang. (2011) Research on single well production prediction based on improved BP neural networks. Science Technology and Engineering. Beijing, vol. 11, pp. 7766–7769.

[7] T. Matsukawa and E. Suzuki. (2016) Person re-identification using CNN features learned from combination of attributes. 2016 23rd International Conference on Pattern Recognition (ICPR), p. 2428-2433.

[8] L. Xing and Y. Qiao. (2016) Deepwriter: a multi-stream deep CNN for text-independent writer identification. 2016 15th International Conference on Frontiers in Handwriting Recognition, p. 584-589.

[9] Y. Bengio, P. Simard and P. Frasconi. (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks. New York, vol. 5, pp. 157–166.

[10] S. Hochreiter and J. Schmidhuber. (1997) Long short-term memory. Neural Computation. Cambridge, vol. 9, pp. 1735–1780.

[11] X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, L. Jiang and Z. Cheng. (2020) Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. Journal of Petroleum Science and Engineering. Amsterdam, vol. 186, Article 106682.

[12] W. Liu, W. D. Liu and J. Gu. (2020) Forecasting oil production using ensemble empirical model decomposition based Long Short-Term Memory neural network. Journal of Petroleum Science and Engineering. Amsterdam, vol. 189, Article 107013.

[13] H. Wang, L. Mu, F. Shi and H. Dou. (2020) Production prediction at ultra-high water cut stage via Recurrent Neural Network. Petroleum Exploration and Development. Beijing, vol. 47, pp. 1009–1015.

[14] Y. Kim, C. Denton, L. Hoang and A. M. Rush. (2017) Structured attention networks. the 5th International Conference on Learning Representations (ICLR 2017), p. 1-21.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin.(2017) Attention is all you need. arXiv, p. 1-15.

[16] D. Jain, A. Kumar and G. Garg. (2020) Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. Applied Soft Computing. Amsterdam, vol. 91, Article 106198.