Research on the forecast of coal price based on LSTM with improved Adam optimizer

Xinrong Liu¹

¹Department of Computer Science and Technology, Shandong Institute of Business and Technology, China

Corresponding author and e-mail: Xinrong Liu, 201511764@sdtbu.edu.cn

Abstract. Coal is a primary basic energy in China and the changes in coal prices will affect the production of its downstream industries. Forecasting coal prices will benefit companies' decision-making on future development. With coal price as the research object, this paper selects 6 factors that affect coal prices, and uses deep learning technology to propose an LSTM neural network model with improved Adam optimizer to forecast it. The experimental results show that the forecast performance of this model is obviously better than that of the traditional model, which not only improves the forecast accuracy, but also provides a reference for the future development and decision-making.

1. Introduction
As an important energy industry, the coal industry is the foundation of the entire national economy of China. The coal consumption every year accounts for about 70% of China's total energy consumption. The research and forecast of the future trend of coal prices seems to be particularly important, which can not only help companies estimate their costs and make judgments, but also lay a foundation for the stable development of the coal market. Deep learning technology has developed rapidly since its introduction, and its nonlinear modeling capabilities for complex data have been widely confirmed. Among them, the neural network algorithm only needs to fit the forecast value according to the input data characteristics, which is a simple mapping between input and output. This mechanism can adapt to the characteristics of the varied price market and complex data structure as well, and is suitable to be applied in the coal price forecast.

Many scholars in China have done research on coal price forecast, the methods of which can be divided into two categories: the first category is the regression method based on the time series model, which uses the historical data of coal prices to study and find its changing laws. Wenqi HE [1] proposed to use ARIMA and linear regression combine model to forecast coal prices. From the perspective of studying the changing laws of coal price sequence itself, Hui NING [2] proposed a forecasting model based on a rolling time window and applied this model into the forecast of the thermal coal price in Qinhuangdao Port. Since it is required that the time series data should be stable when using the time series method to forecast the coal price, while the time series can only capture linear relationships rather than nonlinear relationships essentially, the time series regression model has forecast limitations. Presently, the main price forecast method is of the second category based on AI model. Penglin LI [3] analyzed the factors that affect coal prices based on BP neural network and
believed that the nonlinear mapping relationship of the BP AI neural network model has better processing effects on variables than the linear relationship of the multiple linear regression models. Yang XIAO [4] analyzed the factors that affect the coal prices with the co-integration theory, judged the key factors that affect the coal prices, and established a coal price forecasting model that combined the neural network and the co-integration theory through combining the built-in advantages of the neural network in nonlinear modeling. Over the years, the Long Short-Term Memory (LSTM) neural network in the study of the deep learning method in the AI model has achieved great successes. It has memorability of long short-term information and effectively avoids the problems of vanishing gradient and exploding gradient in the training process of the conventional recursive recurrent neural network models [5] through its own special structural design, and can be trained effectively, thus using historical sequence information. And it is particularly suitable for being used to process time series data [6]. At present, the LSTM technology has been widely studied and applied in many fields. Yongle WANG [7] proposed a coal price model research based on LSTM, indicating that using the LSTM neural network method to study coal prices is a relatively reasonable and advanced method. When using the LSTM method, the Adam optimizer can be used to calculate and update various network parameters that affect model training and output to make them gradually close to and reach the optimal value, thus minimizing or maximizing the loss function to optimize the update iteration of the network and improve the iterative efficiency [8]. However, the basic Adam optimization method in the past can only be based on recent historical gradient values of learning, and the convergence in the later training stage is not sufficient. Therefore, this paper proposes an improved coal price forecast research method based on LSTM with improved Adam optimizer.

2. LSTM model with improved Adam optimization

2.1. Structure of the LSTM model network

Due to the impact of factors such as transportation, supply and demand, and market, and the interaction of various factors, the coal price shows the characteristics of nonlinear changes in the time dimension. The AI neural network can fit the nonlinear mapping and has a relatively strong self-adaptation and self-learning ability. To this end, this paper will adopt an Adam algorithm with a self-adaptive learning rate to improve the training process of the traditional LSTM network to enable the neural network to get a more reasonable network parameter. This model avoids the problem of determining index weights and is completely driven by data. It makes full use of the much stronger nonlinear learning ability of the deep neural network and effectively combines qualitative analysis with quantitative analysis, reduces the impact of human factors, can explore and learn the complex and subtle deep relationship in the inaccurate and noisy effective data set. The model training enables faster convergence and higher forecast accuracy. The LSTM network consists of an input layer, an output layer and a hidden layer. Compared with the traditional recurrent neural network, the unit structure of the hidden layer of the LSTM is relatively special as shown in the figure below:
Figure 1. LSTM model memory cell structure.

Each unit of LSTM has a memory cell which is modified with the introduction of a selective mechanism and can select through the forget gate, the input gate and the output gate respectively. Among them, the forget gate determines what information is discarded by the "cell state", and the information increase gate determines what information to be put in the "cell state", and the sigmoid layer determines what value needs to be updated. The Tanh layer creates a new candidate vector $C_t$ to prepare for the state update. After passing the first and the second gates, the deletion and addition of the transmitted information can be determined, and then the cell state can be updated. Update $C_{t-1}$ to $C_t$ and conduct a dot product between the old state and the output of the forget gate and lose all information that is definitely not needed. The third gate firstly runs a sigmoid layer to determine which part of the cell state is output, and uses Tanh to process the cell state to get a value between -1 to 1, and then conduct a dot product between it and the output of the sigmoid gate to determine the part of the program output. The calculation formula between each gate is as follows:

The forget gate of the first layer:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f)$$  \hspace{1cm} (1)

Wherein $f_t$ represents the output of the forget gate at time $t$, $\sigma$ represents the activation function sigmoid, $w_f$ represents the weight of the forget gate, and $b_f$ represents the bias of the forget gate. $h_{t-1}$ represents the input of the previous moment, and $x_t$ represents the input of the current moment.

The input gate of the second layer:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i)$$  \hspace{1cm} (2)

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t$$  \hspace{1cm} (3)

Wherein $\tilde{c}_t = (\tanh(w_c \cdot [h_{t-1}, x_t] + b_c)$

Wherein $i_t$ represents the output of the sigmoid layer of the input gate at time $t$ and determines what value needs to be updated. $c_{t-1}$ is the previous cell state and is the new candidate vector output from $c_t$ the tanh layer and prepares for state update. $\tilde{c}_t$ is the middle memory and determines which information needs to be enhanced. After the previous cell state passing through the first and the second gates, the deletion and addition of the transmitted information can be determined and the update of the cell state can be conducted.

The output gate of the third layer:

$$o_t = \sigma(w_o [h_{t-1}, x_t] + b_o)$$  \hspace{1cm} (4)

$$h_t = o_t \cdot \tanh(c_t)$$  \hspace{1cm} (5)

Wherein, $o_t$ represents the most original output gate, $h_t$ represents the final output at the current time $t$, which uses the activation function Tanh to process the cell state to get a value between -1 and 1, and then multiplies it with the output of the sigmoid gate to get the determined output. After the data processing of three gates, LSTM establishes a long-term delay relationship between input and output, and the information flow in between the states in the cell structure diagram of the hidden layer is continuous, thus avoiding exploding and vanishing gradients.
2.2. Adam optimization algorithm

In the LSTM learning process, the more the network parameters, the stronger the learning ability. In the training process, there are 4 groups of parameters of forget gate, input gate, output gate and cell state update. Use the training algorithm to get a model that meets the expected effect through constant learning of pervious data. The Adam optimizer has been widely used in the deep learning field by virtue of its self-adaptive learning rate and relatively fast convergence speed. Therefore, this paper adopts the Adam method to update the LSTM model parameters. The core of the Adam method is to dynamically adjust the learning rate of each parameter through calculating the first moment and the second moment of the gradient. Therefore, the parameter update is stable and is not easy to fall into the local optimum. The training steps are described as follows: The first step is to calculate gradient moment estimation at time $t$:

$$m_t = u_1 \times m_{t-1} + (1 - u_1) \times g_t$$  \hspace{1cm} (6)

$$n_t = u_2 \times m_{t-1} + (1 - u_2) \times g_t^2$$  \hspace{1cm} (7)

Wherein, $m_t$ represents the average value of gradient index at time $t$, and $n_t$ represents the squared gradient at time $t$. $m_{t-1}$ represents the average value of gradient index at the previous time, and $n_{t-1}$ represents the squared gradient at the previous time. $u_1 = 0.9$, and $u_2 = 0.999$, These two values are hyper-parameters that control the attenuation of the moving average. The second step is to calculate the update bias: $\hat{m}_t = \frac{m_t}{1 - u_1^t}$, $\hat{n}_t = \frac{n_t}{1 - u_2^t}$, the third step is to get the final value of the parameter update: $\theta_t = \theta_{t-1} - \eta * \frac{\hat{m}_t}{\epsilon + \sqrt{\hat{n}_t}}$, wherein the initial value of the update learning rate is $\eta = 0.01$, and in the sampling training process, it is found that,

The convergence tendency of the model is close to the change characteristics of the power exponential function. Therefore, this paper adds a power exponent correction term to the learning rate at time $t$: $\eta_t = \eta_{t-1} \times \left(1 + \frac{t}{R}\right)^{-\alpha}$

$$\eta_{t-1} = \eta_0 \times \left(1 + \frac{t}{R}\right)^{-\alpha}$$  \hspace{1cm} (8)

$$k = \sum_{i=1}^{n} \lambda_i + q$$  \hspace{1cm} (9)

$$L_t = \varepsilon g_t^2 + g_t^2$$  \hspace{1cm} (10)

Wherein: $R$ is the maximum number of iterations; $\varepsilon$ is the attenuation factor, with the value of 0.99; it can be seen that the improvement of learning rate in this paper is based on the value of the learning rate in the previous stage and the gradient value of the current stage is used for adaptive adjustment. After adding the power exponent correction factor, the convergence of the gradients with the same directions before and after the time $t$ can be accelerated, improving the possibility to avoid the oscillation zone theoretically.

3. Algorithm verification

3.1. Analysis of factors affecting coal prices

The coal price is affected by macroeconomics, supply and demand conditions, logistics costs, and upstream and downstream product prices, and many other factors. Many scholars have analyzed factors that affect it. Based on factor analysis and support vector machine (SVM), Qing XU [9] believed that the main factors that affect coal prices are total coal output, total sales volume, import volume, international crude oil prices and coal prices of the period lagging behind. Yuanqi LIU [10] used a multiple regression linear model to analyze the main factors that affect coal prices, which include inventory index of the Bohai Sea region, key power plant inventories, coastal coal freight index, general steel index and other data. To sum up, the above research authors believed that the main factors that affect coal prices are mainly divided into the following three aspects as show in Table 1 below:
Table 1. Factors affecting coal prices.

| Category                      | Quantitative value                                      | Identifier |
|-------------------------------|--------------------------------------------------------|------------|
| Logistics factors             | Daily average of sea freight                            | X₁         |
|                               | Number of anchorage ships                              | X₂         |
| Supply and demand impact      | Daily average coal inventory of six major power generation groups | X₃         |
|                               | Average daily coal consumption of six major power generation groups | X₄         |
| Market factors                | International coal prices                              | X₅         |
|                               | Coal prices last month                                 | X₆         |

The thermal coal price in the Bohai Sea region has always become a weather vane for changes in coal prices across the country. Therefore, this paper takes coal price as the forecast object and the data of its settlement price as a sample. The sources of the indicator data selected are explained as follows: Since the difference between these 6 characteristic values of daily average of sea freight, number of anchorage ships, coal prices last month, international coal import prices, average inventories of six major power generation groups, and average daily coal consumption of six major power generation groups is relatively big, it is necessary to normalize them, which can enable the data to better adapt to the model and improve its convergence speed and forecast accuracy. The normalization method adopted by this paper is "minimum - maximum normalization" and the calculation formula is as follows:

\[
x^* = \frac{x - \min}{\max - \min}
\]  

The data values are mapped to \([0,1]\) through this method, and the discrete standardization preserves the existing relationship of the original data. This paper selects the thermal coal prices from January 2016 to December 2018 for training, and uses the thermal coal prices from January 2019 to January 2020 as the test set. The experimental programming language is Python3.6 and the development tool is Spyder, with a utilization of the Keras platform, a 2-layer neural network is built and the data are added to the LSTM model for price forecast. The LSTM model in this paper includes 1 input layer and 1 output layer. 2 hidden layers are set. After repeated experiments, 150 neurons are set in each hidden layer, with a total of 300 neurons. According to the deep learning framework of Keras, the optimizer uses the traditional Adam algorithm and the Adam algorithm with the power exponent correction factor to compare the results of the model forecast.

4. Model evaluation

Considering the characteristics and limitations of the evaluation criteria in practical application fields, it is difficult for a single evaluation indicator to fully and comprehensively measure the model training results. This paper adopts three evaluation indicators to test the training effects of the model, namely, MSE (Mean Square Error), MAE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error). The smaller the results of these three values, the smaller the model error. Assuming that the forecast value, the true value \(\hat{y} = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}_n\}\), then \(y = \{y_1, y_2, ..., y_n\}\),

\[
\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} (\hat{y}_t - y_t)^2
\]

\[
\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |\hat{y}_t - y_t|
\]
RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \quad (14)

The forecast and actual comparison obtained and the model error effect diagram are shown as follows:

**Figure 2.** Forecast values of thermal power prices under different models.

**Figure 3.** Error results of Adam optimization and improved Adam optimization models.

5. Conclusions
The forecast of coal prices currently is of great significance for the production guidance of relevant industries. This paper selects six characteristic values from multiple factors that affect coal prices into the forecast model and uses AI method to introduce the LSTM forecast method with improved Adam optimizer, and introduces the power exponent correction into the model training. The experimental study shows that after the introduction of the new algorithm, compared with the traditional Adam optimization method, the coal price forecast in the Bohai Sea region is better with reduced model training errors and improved accuracy.

References
[1] Wenqi HE. Coal Price Forecast Based on ARIMA and Linear Regression Combined Model. I. China Management Informationization, 2021 (1): 187-188
[2] Hui NING. Research on Coal Price Prediction Model Using the ε-SVR Method Based on Sliding Time Window J. Coal Economic Research, 2020 (3): 12 - 18.
[3] Penglin LI. Research on the Influence Price Based on Factors and Prediction of Coal BP Neural Network. J. Mathematics in Practice and Theory, 2015 (9): 113-126
[4] Yang XIAO. The Coal Price Forecasting Model Based on the Co-integration Theory and ANN D. Xi'an University of Science and Technology. 2016
[5] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [M]/Supervised Sequence Labelling with Recurrent Neural Networks. Springer Berlin Heidelberg, 1997: 1735-1780.
[6] DONAHUE J, et al. Long-term recurrent convolutional networks for visual recognition and description [EB/OL] 2020-03-10 https://arxiv.org/abs/1411.4389.
[7] Yongle WANG. Research on Coal Price Forecasting Model Based on LSTM D. Xi'an University of Science and Technology. 2020
[8] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization C //Proc of the 3rd Int Conf for Learning Representations. US San Diego: ICLR, 2015
[9] Qing XU. Research on the Coal Price Forecast of Qinhuangdao - Based on Factor Analysis and SVM. J Price: Theory & Practice 2014.02 79-81
[10] Yuanqi LIU. Research on Spot Price Forecasting Model of Thermal Coal in the Bohai Rim J Coal Economic Research 2020.10 19-22