An Improved CNN–LSTM Model for Daily Gold Price Prediction

Yi Lin

Detroit Green Technology Institute, Hubei University of Technology, Wuhan, Hubei, China
rickylam1212@163.com

Abstract. Gold prices play a pivotal role in the global financial system, but they often experience short-term volatility, long-term trends, and frequent regime switching. To address this, this paper proposes a new architecture, building on the CNN architecture and combining ordinary convolution with dilated convolution to create a parallel multi-scale convolution. This architecture then utilizes a learnable gating coefficient, which is then dot-producted with two branches before entering the LSTM. Ablation and comparative experiments were conducted with two training sets: test set ratios of 6:4 and 8:2. Experimental results show that, in the 6:4 ratio, the parallel dilated convolution without gates (DilatedCNN-LSTMs) performs best, achieving an RMSE of 13.11 and a MAE of 8.89. In the 8:2 ratio, the parallel model with gates (DilatedCNN-LSTMs) achieves the best RMSE of 15.31 and a MAE of 10.22. When the training set ratio is large, gating can effectively reduce bias and improve generalization. However, when the training set ratio is small, the increased degrees of freedom can lead to variance and training instability. This method requires no exogenous variables, is low-cost, and highly reproducible, making it valuable for daily gold price forecasting and risk management.

Keywords: Gold Price, Deep Learning, LSTM, CNN.

1 Introduction

Gold plays an important role in the market. With the global monetary policy, geopolitics, and various uncertainties increasing, the daily gold price is often accompanied by complex features such as nonlinearity, regime switching, and long memory [1]. At the same time, from the perspective of investors and policymakers, gold is often used to hedge against inflation, avoid economic risks, for asset allocation, and policy evaluation [2]. Therefore, this places high demands on the accuracy and generalization ability of short- and medium-term forecasts. Traditional statistical models, such as ARIMA [3] and various extensions, are relatively strong in characterizing linear dependencies, but the instability and mutation characteristics of daily gold prices make ARIMA not advantageous. Machine learning methods, such as Annas et al. using SVR to predict Indonesian gold prices [4], and Pierdzioch et al. using random forest methods to predict precious metal prices [5]. Although these traditional machine learning methods are better at characterizing nonlinear features and can better fit, they still have difficulty in taking into account local features and cross-window time

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_53

series features. With the development of the times, various deep learning methods have emerged, such as CNN, which is good at robustly capturing features from local windows [6], and LSTM, which is better at modeling cross-window dependencies [7]. Meanwhile, Santika[8] proposed a CNN+LSTM approach to take into account both local and cross-window features, which improved model performance. However, the CNN in this model structure is limited by its small receptive field[9] and is not adaptable enough to price fluctuations caused by factors such as regime switching. Therefore, this paper proposes a new architecture that maintains the original CNN branch while adding another dilated convolution branch. At the same time, when the training set sample size is large, the weights of the two branches are adaptively controlled by learning coefficient gating, and then input into the LSTM to aggregate temporal dependencies. With only a small number of parameters added, the performance of the model is improved efficiently and lightly, allowing the new model to better cope with the complex characteristics of daily gold prices and achieve better predictions.

2 Dataset Preprocessing

This study used a dataset of 4,175 daily univariate gold price series covering the period 2005-2021, all denominated in US dollars (USD). The data is sourced from the World Gold Council's official website. The Council uses a combination of London Bullion Market Association (LBMA) pricing and several regional quotes. The LBMA price has long served as the core benchmark for global gold pricing, while regional quotes are important indicators for local markets. This study employed two training and test set partitioning schemes: an 8:2 and a 6:4 ratio, respectively. The 8:2 ratio evaluates the potential performance of different models on a larger training set and prevents models requiring large training data from underperforming due to limited data. The 6:4 ratio better tests whether the model is robust to distributional drift, enhances the confidence of statistical inferences, and provides a more rigorous assessment of model performance.

To achieve better training results, this study used data normalization in both the training and test sets to address the issue of large differences in numerical ranges. The experiment used Min–Max scaling as shown in formula (1). x represents the data to be normalized, max(x) represents the maximum value, and min(x) represents the minimum value. The final result then undergoes a shape transformation before entering the model designed for this study, which contains both input and output data. Specifically, four consecutive rows of gold prices are used as input, and the first row after that is used as output.

$$\hat{x} = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

3 Model Optimization

3.1 Model Nnetwork Architecture

This study combined CNNs and LSTMs, but the direct concatenation of CNNs and LSTMs performed poorly. This was primarily due to the limited receptive field of the CNN stage, which was unable to capture long-term memories, slow trends, and regime switching. Consequently, information in the front-end could not be fully encoded by the CNN branch. Once the information was entered into the LSTM, the LSTM was forced to compensate for the lost context over long timescales. On the one hand, the need to accumulate evidence across multiple steps to identify long-term dependencies and regime states increased the variance of parameter estimates and made training more unstable. On the other hand, despite the LSTM's gating mechanism, which helps mitigate vanishing gradients, information still fades over time, increasing overfitting and computational cost. Furthermore, the front-end processing did not explicitly separate structural components from high-frequency noise, resulting in a low signal-to-noise ratio for the representation entering the LSTM. This made the gating more susceptible to short-term perturbations, making it less effective in capturing slow trends and overly sensitive to noise or local features. When the gold price changes its characteristics due to a regime switch, the LSTM requires multiple time steps to learn its internal features. During this transition period, the model still uses the representations and states of the old regime to predict data for the new regime. As a result, the instantaneous residual increases sharply near the switch point (which is particularly sensitive to squared losses such as RMSE), resulting in error spikes. This is especially true in the 6:4 dataset scheme, where the evaluation period often spans multiple switches. The accumulation of these spikes can degrade long-window metrics and test results. Therefore, this paper proposes a new structure, as shown in Figure 1.

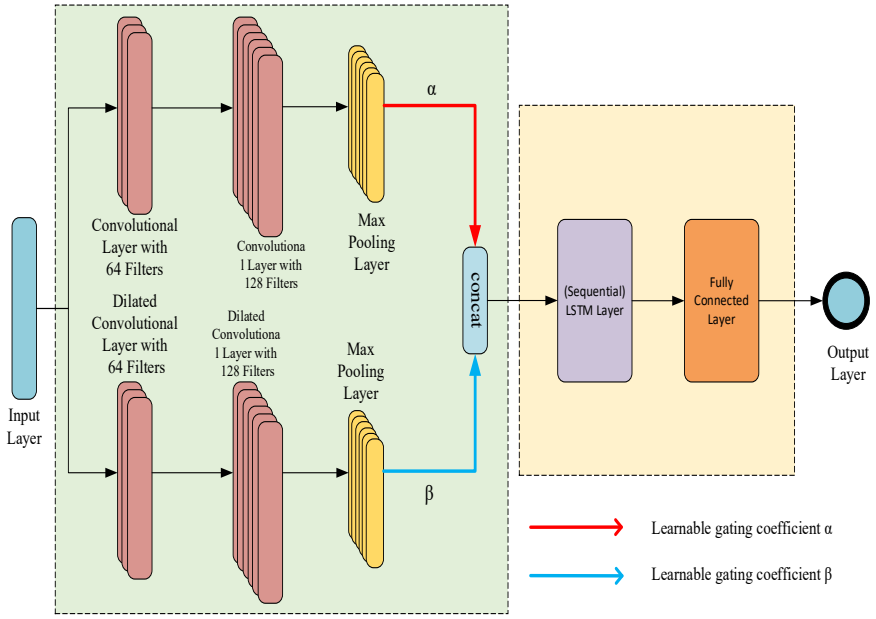

Fig. 1. Improved CNN+LSTM model architecture (Picture credit: Original).

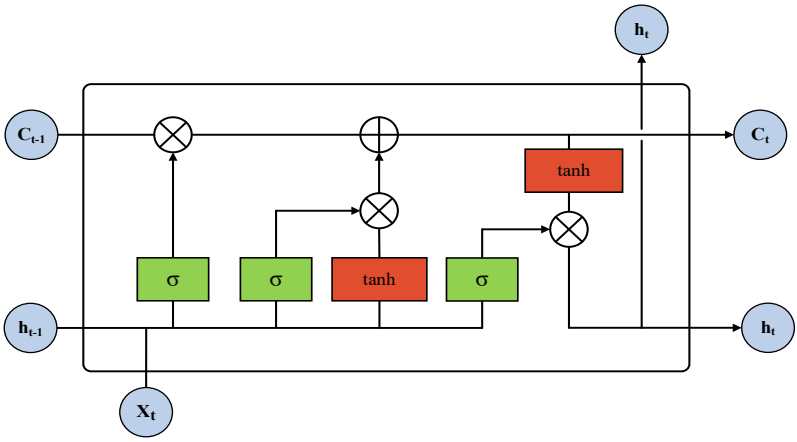

Fig. 2. LSTM principle diagram.

As shown in formulas (2) to (15) below, a new network architecture is obtained after combining the dilated convolution branches and the learnable gating coefficients. After

the new parallel branch processing, compared with the original structure, when entering the LSTM structure shown in Figure 2, the burden of LSTM "compensating" for front-end information over long spans is reduced, reducing parameter estimation variance and training instability; when there are sufficient samples, compared with simple unweighted splicing, the introduction of gating can suppress redundant or mutually interfering branch responses, improve the signal-to-noise ratio, and enhance interpretability. This study calls the model with a parallel expansion rate of 1, that is, the ordinary convolution branch and the fusion of learnable gating coefficients DualCNN-LSTM, and the model without the fusion of learnable gating coefficients DualCNN-LSTMs; the model with a dilated convolution parallel branch with a expansion rate of 2 and the fusion of learnable gating coefficients DilatedCNN-LSTM, and the model without the fusion of learnable gating coefficients DilatedCNN-LSTMs.

3.2 Parallel Multi-Scale Convolution

While retaining the original structure, this study adds another dilated convolution [10] branch to capture short-term local features. The parallel branch uses dilated convolution with the same kernel length and a fixed dilation rate of 2. While adding a small number of parameters, it effectively expands the effective receptive field and solves the problem of difficulty in capturing long-term memory, slow dynamics, and regime switching features. The specific processes of these two convolutions are shown in formulas (2)-(9).

Ordinary convolution branch:

$$X_{1,1} = TimeDistributed\big(Conv1D(X;\ f = 64, k = 2, \sigma = ReLU, p = same)\big) \quad (2)$$

$$X_{1,2} = TimeDistributed\Big(Conv1D\big(X_{1,1};\ f = 128, k = 2, p = same\big)\Big) \quad (3)$$

$$X_{1,3} = TimeDistributed\Big(MaxPool1D\big(X_{1,2};\ s = 2\big)\Big) \quad (4)$$

$$X_1 = TimeDistributed\Big(Flatten\big(X_{1,3}\big)\Big) \in R^{B\times T\times F_1} \quad (5)$$

Dilated convolution branch:

$$X_{2,1} = TimeDistributed\big(DilatedConv1D(X;\ f = 64, k = 2, \sigma = ReLU, p = same, d = dilation)\big) \quad (6)$$

$$X_{2,2} = TimeDistributed\Big(DilatedConv1D\big(X_{2,1};\ f = 128, k = 2, p = same, d = dilation\big)\Big) \quad (7)$$

$$X_{2,3} = TimeDistributed\Big(MaxPool1D\big(X_{2,2};\ s = 2\big)\Big) \quad (8)$$

$$X_2 = TimeDistributed\Big(Flatten\big(X_{2,3}\big)\Big) \in R^{B\times T\times F_2} \quad (9)$$

The dilation is set to 2 in all experiments in this study.

3.3 Learnable Gating Coefficients

The simple accumulation of two branches will bring redundant noise, so the outputs of the two branches are multiplied by the learnable gating coefficients [11] respectively,

which can dynamically adjust the contribution of the ordinary convolution and dilated convolution branches as the fluctuation regime changes. Especially when there are sufficient samples, that is, when the training set is large enough, because the samples cover more regime switches or extreme cases, this mechanism will make the model more stable. However, when the sample size is small, it is more susceptible to noise [12], causing the coefficients to sometimes lean towards the ordinary convolution branch and sometimes towards the dilated convolution branch, resulting in a decrease in generalization ability. The calculation process of the learnable coefficient gating is shown in Formulas (10) and (11).

$$\widehat{X_1} = \alpha \cdot X_1, \alpha \geq 0 \tag{10}$$

$$\widehat{X_2} = \beta \cdot X_2, \beta \geq 0 \tag{11}$$

4 Experimental Analysis

4.1 Evaluation Indicators

The evaluation indicators used in this paper are RMSE, MAE, and R2, and the calculation process is as follows. N: The number of samples (test set size), y_i: The true value of the i-th sample, $\hat{y}_i$: The predicted value of the i-th sample, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$: The mean of the true values in the test set.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

RMSE indicates the magnitude of the overall forecast error and is more sensitive to large errors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{13}$$

MAE represents the absolute magnitude of the mean error and characterizes the typical error level. Compared to RMSE, it focuses more on the deviation of the overall median level.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{14}$$

The R2 display model is compatible with the level of comprehensibility/comprehension of the model.

4.2 Ablation Experiment

Table 1. The ratio of the training set and test set is 6:4.

	RMSE	MAE	R2
CNN-LSTM	15.12	10.52	0.9983
DualCNN-LSTM	13.40	13.40	0.9985
DualCNN-LSTMs	13.23	8.96	0.9986
DilatedCNN-LSTM	13.12	8.89	0.9987
DilatedCNN-LSTMs	13.11	8.89	0.9987

Table 2. The ratio of the training set and test set is 8:2.

	RMSE	MAE	R2
CNN-LSTM	16.27	11.14	0.9977
DualCNN-LSTM	15.67	10.55	0.9981
DualCNN-LSTMs	15.86	10.73	0.9980
DilatedCNN-LSTMs	15.36	10.25	0.9983
DilatedCNN-LSTM	15.31	10.22	0.9983

As shown in Tables 1 and 2, all models converged after training. After incorporating the parallel convolutional structure, performance significantly improved compared to CNN-LSTM. After incorporating dilated convolutions, it is clear that both DilatedCNN-LSTM and DilatedCNN-LSTMs outperform both standard convolution and standard convolution plus parallel structures, regardless of dataset size.

Figure 3 shows the prediction results for a training set to test set ratio of 6:4. DilatedCNN-LSTMs perform better without the learnable gating coefficients. This indicates that, when the training set is small, adding this learnable gating coefficient increases the model's degrees of freedom, raising the model's variance and causing significant fluctuations in the gating parameters. The multiplicative structure of the gating can lead to over-suppression of early branch gradients, resulting in "dead branches" and unstable convergence. Furthermore, gating and LSTM or normalized scaling present discriminability issues. At the same time, dilated convolutions already provide significant multi-scale benefits and diminishing returns. Therefore, DilatedCNN-LSTMs perform better and generalize better when the training set to test set ratio is 6:4.

Figure 4 shows the prediction results when the training set to test set ratio is 8:2. Larger training set sizes are supported by dilated convolutions, a parallel structure of conventional convolutions, and an adaptive fusion approach. The parallel structure of DilatedCNN-LSTM captures features of different lengths and then performs a properly weighted concatenation. This maintains multi-scale feature extraction while reducing redundancy and preventing poorly performing branches from taking on high weights. When the training set is further restructured, the approximation error caused by gating decreases, meaning the bias decreases, offsetting the cost of the additional parameters caused by adding gating, resulting in better performance on the test set.

4.3 Comparative Experiment

Table 3. The ratio of training set and test set is 6:4

	RMSE	MAE	R2
CNN	13.61	9.38	0.9986
LSTM	30.24	26.00	0.9925
RandomForest	14.23	10.02	0.9998
DilatedCNN-LSTMs	13.11	8.89	0.9987

Table 4. The ratio of training set and test set is 8:2

	RMSE	MAE	R2
CNN	16.99	11.32	0.9982
LSTM	15.75	10.52	0.9984
RandomForest	16.03	10.74	0.9998
DilatedCNN-LSTM	15.31	10.22	0.9983

As shown in Tables 3 and 4, all models achieved convergence. The results consistently demonstrate that the models incorporating parallel dilated and normal convolutions outperform traditional machine learning methods like RandomForest, deep learning methods like CNN, and LSTM, achieving superior performance in terms of both RMSE and MAE. When the sample size is more plentiful, such as when the training set to test set ratio is 8:2, a hybrid structure combining gated concatenation of learnable coefficients followed by LSTM performs best. Specifically, with a 6:4 dataset ratio, the DilatedCNN-LSTMs achieved RMSE and MAE of 13.11 and 8.89, respectively, representing improvements of at least 3.67% and 5.22% compared to the other methods. With an 8:2 dataset ratio, the DilatedCNN-LSTMs achieved RMSE and MAE of 15.31 and 10.22, respectively, representing improvements of at least 2.79% and 2.85% compared to the other three methods. This demonstrates the effectiveness of this study's improvements and demonstrates promising results for daily gold price forecasting.

4.4 Visualization of Prediction Results

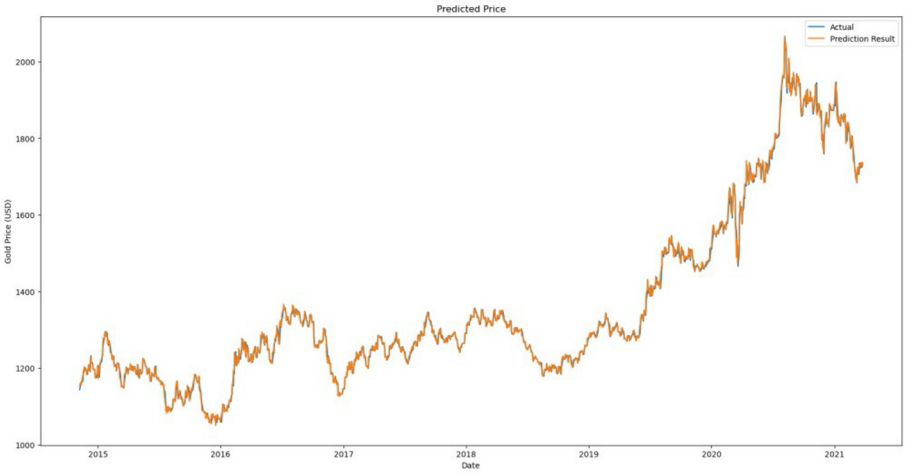

Fig. 3. DilatedCNN-LSTMs prediction results (training set to test set ratio 6:4) (Picture credit: Original)

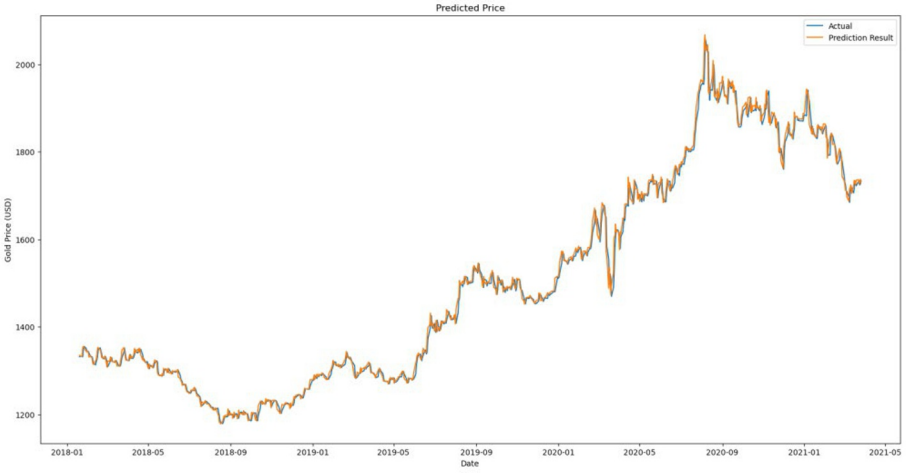

Fig. 4. DilatedCNN-LSTM prediction results (training set to test set ratio 8:2) (Picture credit: Original).

As can be seen from Figures 3 and 4, the improved new model performs well in daily gold price forecasting, especially when the DilatedCNN-LSTMs training set and test set ratio is 6:4. The prediction curve can respond to fluctuations more promptly and maintain a good fitting effect.

5 Conclusion

This study uses 4,175 daily univariate gold prices from 2005 to 2021 as the research object. Based on improvements to CNN+LSTM, this paper proposes an improved model combining parallel multi-scale (normal convolution and dilated convolution) with learnable gating and LSTM. This study conduct ablation and comparative experiments using two datasets with different training and test set ratios: 6:4 and 8:2, respectively. The results demonstrate the effectiveness of parallel multi-scale. Introducing this parallel structure yields improvements in both dataset configurations, outperforming models such as CNN and RandomForest. The dataset configuration influences the benefits of learnable gating. At a 6:4 ratio, the parallel dilated convolution without gates (DilatedCNN-LSTMs) achieves the best RMSE of 13.11 and MAE of 8.89. At an 8:2 ratio, the parallel model with gates (DilatedCNN-LSTMs) achieves the best RMSE of 15.31 and MAE of 10.22. The results show that adaptive fusion can effectively reduce bias and improve generalization when samples cover more regimes and extreme cases. Otherwise, it increases degrees of freedom and leads to unstable training. This low-cost and efficient architecture, using dilated convolutions to expand the receptive field, parallel processing to retain multi-scale information, gating for conditional reweighting, and LSTM to incorporate long-term memory, can achieve significant performance improvements without introducing exogenous variables. This has significant practical implications for daily gold price forecasting and risk management.

References

1. Arouri, M. E. H., Hammoudeh, S., Lahiani, A., Nguyen, D. K.: Long memory and structural breaks in modeling the return and volatility dynamics of precious metals. The Quarterly Review of Economics and Finance 52(2), 207–218 (2012)
2. Zulaica, O.: What share for gold? On the interaction of gold and foreign exchange reserve returns. Bank for International Settlements, No. 906 (2020)
3. Jenkins, G. M.: Autoregressive–Integrated Moving Average (ARIMA) Models. In: Encyclopedia of Statistical Sciences, vol. 1, pp. 1–??. Wiley, New York (2004)
4. Annas, S., Rais, Z., Aswi, A.: Implementation of Support Vector Regression (SVR) Analysis in Predicting Gold Prices in Indonesia. In: 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023), pp. 97–107. Atlantis Press, Paris (2023)
5. Pierdzioch, C., Risse, M.: Forecasting precious metal returns with multivariate random forests. Empirical Economics 58(3), 1167–1184 (2020)
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (2002)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997)
8. Santika, I. W. K. G., Sa'adah, S., Yunanto, P. E.: Gold price prediction using convolutional neural network-long short-term memory (CNN-LSTM). Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control (2021)

9. Luo, W., Li, Y., Urtasun, R., Zemel, R.: Understanding the effective receptive field in deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 29, pp. 1–??. Curran Associates, Red Hook (2016)
10. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
11. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. IEEE, New York (2018)
12. Hastie, T., Tibshirani, R., Friedman, J., Franklin, J.: The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 27(2), 83–85 (2005)