Research on an Improved Transformer Model for Predicting Crude Oil Prices

Heng Guan

1324499282@qq.com

Chongqing University of Posts and Telecommunications Software Engineering Instituteline Chongqing, China

Abstract. To enhance the precision of crude oil price prediction, this study introduces an innovative method that integrates the attention mechanism from transformer models with normal distribution functions. The incorporation of normal distribution functions aids in capturing the inherent volatility within each segment of crude oil price data, thereby preserving the distinctive characteristics of historical data. This preservation is instrumental in achieving more accurate predictions of future crude oil prices, consequently facilitating more reasoned projections of crude oil price trends. Our investigation is centered on the daily price data of West Texas light crude oil spanning from January 1, 2001, to January 1, 2023. Subsequently, an improved transformer model was employed to train and predict the aforementioned dataset. Comparative analysis against the benchmark model reveals the superior predictive performance of the enhanced transformer model in comparison to traditional transformer models and LSTM models. Moreover, the research results have successfully withstood rigorous robustness testing, affirming the reliability of the proposed model.

Keywords-transformer models;normal distribution functions;crude oil price prediction;

1. Introduction

As a vital chemical raw material and one of the world's largest traded commodities, the price fluctuations of crude oil exert a profound impact on the stability of the global economy, financial markets, and various regions. Positioned as the world's largest developing country, China's relentless surge in crude oil demand is noteworthy. According to BP's 2022 World Energy Statistical Yearbook, China consistently holds the top rank globally in annual crude oil consumption, vividly underscoring its substantial reliance on foreign crude oil[1]. Consequently, the accurate prediction and pricing of international crude oil prices have evolved into fundamental concerns for China's energy supply and demand security, as well as economic stability.

Precisely forecasting crude oil prices and discerning their trends has become a strategic objective for government de- partments, enterprises, and other investors, aiming to for- mulate more scientifically grounded institutional strategies, production investments, and risk mitigation plans. However, in contrast to ordinary commodities, crude oil exhibits intricate properties, thereby designating the international crude oil market as a typical complex nonlinear system.

MSIEID 2023, December 08-10, Guangzhou, People's Republic of China

DOI 10.4108/eai.8-12-2023.2344782

2.Ease of Use

One of the paramount benchmark oils in the international crude oil market is West Texas Medium crude oil. Analysis of the fluctuations in the benchmark oil price reveals that changes in crude oil prices stem from a complex interplay of various factors. Currently, the predominant methods can be broadly categorized into three groups. The first category encompasses statistical prediction methods, exemplified by the ARIMA model[2]. These models leverage mathematical formulas to derive the linear relationships within WTI crude oil price data but fall short in capturing the nonlinear dynamics of crude oil prices.

The second category comprises machine learning-based prediction methods, predominantly featuring support vector machine models and decision tree models. However, these models necessitate intricate feature engineering and exhibit a high dependence on training data, rendering them less adept at handling extensive datasets.

The third category focuses on prediction methods based on deep learning, prominently incorporating recurrent neural networks (RNNs) [3]. Despite advancements like LSTM[4] and GRU[5], experiments[6] have underscored limitations in effectively capturing long-term dependencies, with the LSTM model demonstrating a discernment capability limited to approximately 50 nearby labels out of a context size of 200 labels.

Crucially, the prediction of crude oil price data often en- tails both long-term and short-term repetitive patterns, with modeling long-term dependencies emerging as a pivotal factor in enhancing model performance.Recently, the Transformer has emerged as a groundbreaking time series prediction ar- chitecture, leveraging attention mechanisms to process diverse data sets. The Transformer model's distinctive feature is its ability to access any historical data section, regardless of distance, rendering it particularly adept at capturing circu- lar patterns with long-term dependencies in crude oil price data.However, the self-attention mechanism in this model matches insensitive data in queries and local data, leading to information loss during the prediction process and posing an optimization challenge.In response to the above-mentioned issues, our contributions mainly include the following two points:

- We have successfully applied the Transformer model architecture to the problem of crude oil price prediction. We used technical indicator formulas to extract data features related to the financial attributes of crude oil price data from the raw data, and conducted extensive experiments using the generated data features to verify the potential value of technical indicators in extracting financial related relationships from crude oil price data.
- We add a normal distribution function to the self attention mechanism, calculate the normal distribution for each segment of data, and then input it into the self attention mechanism layer for calculation, in order to retain the non-stationary information of the data and improve the prediction accuracy of the model.

3.Related work

A key challenge in current domestic and international research on crude oil price prediction is that the obtained features cannot fully reflect the value of crude oil, which largely leads to low accuracy in crude oil price prediction. The inherent characteristics of crude oil price data, including nonlinearity, complexity, non stationarity, asymmetry, and long memory, pose significant obstacles to achieving accurate predictions. Dealing with these challenges is complex and arduous.The exploration of the internal influencing factors of crude oil reveals four key aspects: inventory, demand, supply, and technical indicators. It is worth noting that technical indicators are the key decision-making criteria for quantitative investment and play a crucial role in trading operations. Technical analysis of oil prices can reflect changes in market investor sentiment and convey implicit information about asset price fluctuations. Three commonly used technical analysis methods include moving average method, momentum method, and fluctuation method.In addition, to address the challenge of modeling long-term correlations in crude oil price data, a Transformer model architecture was adopted. This facilitates the extraction of cyclic patterns associated with long-term dependencies in the data. In addition, a normal distribution function is added to the self attention mechanism layer to preserve non-stationary information and improve the prediction accuracy of the model.

3.1 Technical Indicators

Technical indicators are powerful tools for analyzing and predicting crude oil prices. They excel at tracking trends, fil- tering out noise in data, responding to market changes quickly,

and uncovering underlying factors influencing price shifts. This study focuses on three key types of indicators—trend, momentum, and volatility—to address the challenges of un- certainty, instability, asymmetry, nonlinearity, and long-term memory in crude oil price data. To enhance the crude oil prediction model, efficiently capturing nonlinear dynamics in time series for optimal performance, we utilize trend indicators such as SMA (Simple Moving Average), EMA (Exponential Moving Average), and KAMA (Kaufman's Adaptive Moving Average). Additionally, momentum oscillators like CCI (Commodity Channel Index) and fluctuation indicators such as ATR (Average True Range) are employed to extract key features.

Moving Average (MA) is a widely used method for analyz- ing data points by calculating average values within different subsets of time series. SMA represents a simple moving average, EMA stands for exponential moving average, and KAMA is the Kaufman adaptive moving average. The specific calculation formula is as follows:

- Simple moving average calculation formula:

$$SMA_t = \frac{1}{T}\sum_{i=k-T+1}^{k} C_i \qquad (1)$$

- The formula for calculating the exponential moving average is:

$$EMA_t = C_{i-1} * p - EMA_{t-1} * (p-1) \qquad (2)$$

- Kuffman Adaptive Moving Average Formula:

$$\begin{cases} KAMA_t = KAMA_{t-1} + S_t * (C_i - KAMA_{t-1}) \\ S_t = [R_t * (f - s) + f]^2 \\ Rt = \frac{|C_i - C_{i-T}|}{\sum_{j=t-T+1}^{t} |C_i - C_{i-j}|} \end{cases} \qquad (3)$$

Ci represents the closing price of crude oil, T represents the interval between cycles, and p represents the index value.The momentum indicator is used to determine when investment instruments have reached an oversold or overbought state, and its main purpose is to detect the beginning and end of trends in the time series.

- The formula for calculating the momentum indicator CCI is as follows:

$$\begin{cases} CCI = \frac{TP_t - SMA_t}{y * MD_t} \\ MD_t = \sum_{i=1}^{T_{CCI}} \frac{\left|TP_{i-T_{CCI}+1} - SMA_{i-T_{CCI}+1}\right|}{T_{CCI}} \\ TP_t = \frac{H_t + L_t + C_t}{3} \end{cases} \tag{4}$$

CCI represents a specific cycle value, while Ht, Lt, and Ct represent the highest, lowest, and closing prices, respectively. The volatility index is used to measure the volatility in crude oil prices, mainly including the ATR index, which does not determine the direction of prices, but can provide important information on price differences and price fluctuations caused by many factors.

- The calculation formula for ATR indicators is as follows:

$$\begin{cases} ATR_t = \frac{1}{T_{ATR}} \sum_{i=1}^{T_{ATR}} TR_{i-T_{ATR}+1} \\ TR_t = \max(|H_t - L_t|, |H_t - C_{t-1}|, |L_t - C_{t-1}|) \end{cases} \tag{5}$$

Among them, Ht, Lt, and Ct are the highest, lowest, and closing prices, respectively, and T is the size of the selected time period. For historical data on crude oil prices, feature extraction is performed using the above formula to obtain an alternative data feature set.

3.2 Transformer model

The Transformer model was originally designed for sequence to sequence tasks such as machine translation [7,9], and with the development of this field, it has evolved to find applications in time series prediction. This model architecture is suitable for contemporary neural networks, which accept a sequence of input data and transform it into another sequence - a process that is particularly effective in capturing the entire input sequence information and generating output sequences through transformation.

The Transformer model is rooted in the concept of attention mechanisms and inspired by human visual attention regulation mechanisms, allowing networks to focus on specific aspects of input sequences during encoding and decoding processes. This method significantly improves decoding efficiency. Compared to LSTM models, Transformer excels at using distributed GPUs for parallel training, processing lengthy texts, and capturing extended semantic correlations.

Despite these advantages, the Transformer model has a tendency to flatten time series data in its self attention mechanism, which may lead to information loss of the original data. Therefore, unlike traditional methods, while using self attention mechanisms, special attention is paid to the inherent non stationarity within the original sequence. This ensures the preservation of non-stationary information present in the original data, as shown in the model structure diagram in **Figure 1**. The non-stationary attention mechanism module is used to capture non-stationary information.

3.3 Non-stationary Attention

In a typical self-attention mechanism, the model generates the same stationary input mean and variance for each seg- ment of time series data. This uniform treatment may cause the model to allocate identical attention, overlooking certain information associated with non-stationarity. This results in the issue of excessive stationarity. To address this problem stemming from the self-attention mechanism calculations, non- stationary factors are introduced concurrently with inputting time series data. These factors are integrated to replace the non-stationary data information, mitigating the impact of over- stationarity in the model.

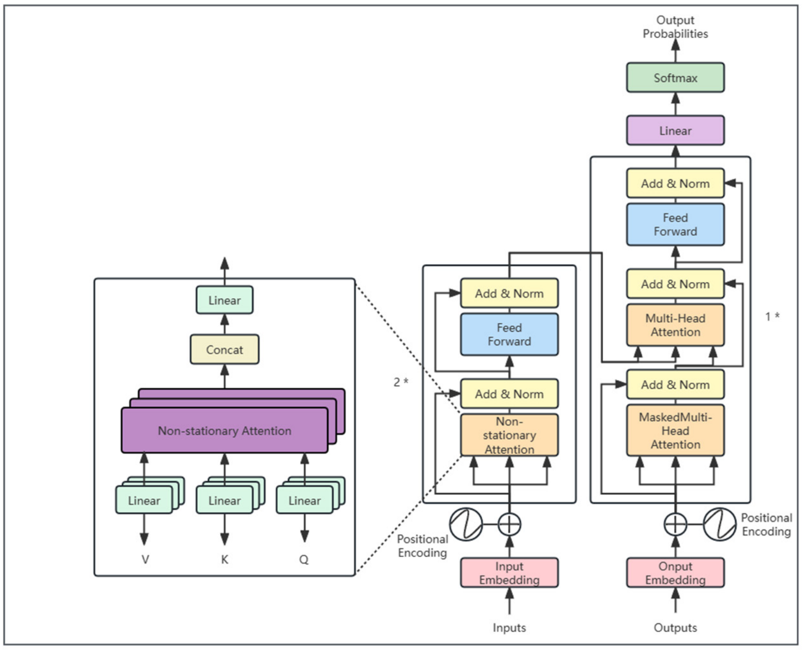

Fig. 1. From the structure diagram of the Transformer, it can be seen that the model mainly relies on the self attention mechanism to calculate the input and output models of the network structure. For the self attention mechanism, it generates the query matrix Q, key matrix K, and value matrix V from the input vector of each encoder. In this study, the input part of the self attention mechanism is modified by adding non-stationary factors w1 and w2, Optimizing parameters through linear layers to obtain the optimal non- stationary factor.

self-attention The main form of the calculation formula for the general self attention mechanism is as follows:

$$Self-attention = Soft\max(\frac{QK^T}{\sqrt{d_k}}V) \tag{6}$$

Here, dk is the size of the input dimension. However, due to the addition of attention mechanism, the order constraint of the Transformer on the original sequence of the sequence disappears, resulting in the loss of order based information in the original sequence. To solve this problem, the Transformer also encodes the positional information in the original sequence, which is called positional encoding. The mathematical expression for positional encoding is as follows:

$$\text{Position}(even) = \sin(\frac{pos}{10000^{\frac{2*i}{d_{model}}}}) \tag{7}$$

$$Position(odd) = \cos(\frac{pos}{10000^{\frac{2*i}{d_{model}}}}) \tag{8}$$

Here, pos is the position of the original data in the sequence.When it is in an even position, sine is used for position encoding. When it is in an odd position, cosine is used for position encoding. i is the sequence number of the original data's position in the sequence, and dmodel is the dimension used by the model. The position encoding is mainly based on trigonometric functions and difference product formulas, which can express position information. We will input $x=[x1,x2, x3, x4,..., xs]^T$. After positional encoding, we can obtain $Q=[q1, q2, q3,..., qs]^T$. Assuming that the mean of Q is δ and the variance is μ, then the Q and K of each input data segment can be transformed into general attention mechanisms as follows:

$$Soft\max(\frac{QK^T}{\sqrt{d_k}}) = Soft\max(\frac{\delta_x^2 * Q' * K'^T + \mu_Q^T * K^T + Q * \mu_K - (\mu_Q^T * \mu_K)}{\sqrt{d_k}}) \tag{9}$$

Simplify the above formula:

$$Soft\max(\frac{QK^T}{\sqrt{d_k}}) = Soft\max(\frac{\delta_x^2 * Q' * K'^T + \mu_Q^T * K^T}{\sqrt{d_k}}) \tag{10}$$

Because the mean δx and variance μQ of each input data segment, which means that the non-stationary factor is not a scalar, we use relative values for estimation and two linear layers to estimate the value of the non-stationary factor.

$$\log w1 = Linear(\delta_x, x), w2 = Linear(\mu_x, x) \tag{11}$$

4.Experiment

We conducted comparative experiments with commonly used models to evaluate the performance of the proposed method on crude oil price prediction benchmarks and further validate its effectiveness. The overall experimental flowchart is shown in **Figure 2**.

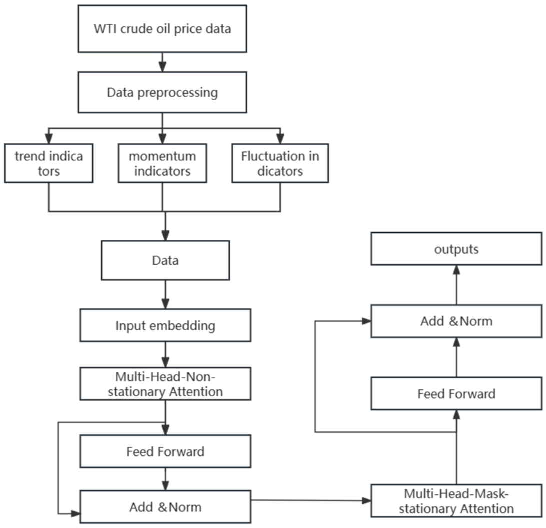

Fig. 2. This flowchart describes the overall process of this research experiment

4.1 Data

The dataset selected the WTI international crude oil trading data from 2001 to 2023 as the basic dataset, and **Figure 3** shows the WTI crude oil price data with daily intervals. This dataset has 13 data features collected at time intervals of day, hour, and week. Non stationarity testing was conducted on this dataset.

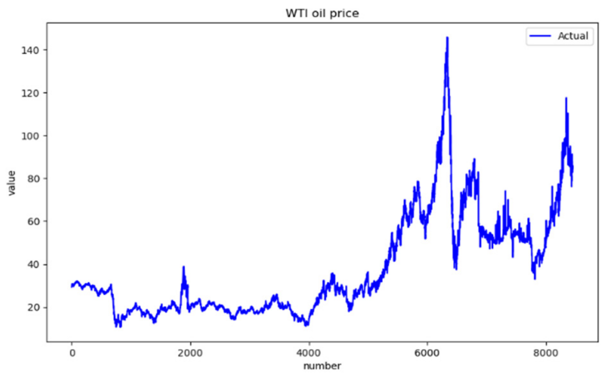

Fig. 3. Data trend chart

Table 1.Data Analysis

Dataset	***Variable Number***	***Frequency***	***Total***	***ADF***
WTI oil	16	1 Hour	12000	-10.32
WTI oil	16	1 Day	8400	-2.23
WTI oil	16	1 Week	1200	-1.46

In the data analysis stage, the enhanced Dickey Fuller (ADF) test statistic [12] was used to evaluate crude oil price data from different intervals. The evaluation results are shown in **Table 1**. The application of ADF test aims to evaluate the inherent non-stationary characteristics of crude oil prices, affirming the theoretical rationality of research methods in solving non-stationary problems. This rigorous testing process enhances the credibility and effectiveness of our approach to dealing with the unique challenges posed by the non-stationary nature of crude oil price data. Next, we plan to explore more statistical tests and analyses to further validate the robustness of our research methods.

4.2 Indicator results

This experiment comprehensively compared the Arima model [2], Xgboost model [10], LSTM [4,11] model, and NS Transformer [8] model without including additional data features, thus verifying the effectiveness of the proposed method. The key indicators evaluated in the experiment include RMSE, MSE, MAE, and R2. The values of these indicators are shown in **Table 2**, and the experimental results show that the predictive performance achieved in this study is superior to other models. This strong comparison highlights the effectiveness of our method of predicting crude oil prices. Looking ahead, our goal is to explore additional technical indicators and refining methods in greater depth to further enhance the predictive ability and applicability of our model, thereby addressing the inherent complexity in crude oil price data.

Table 2. Indicator Results

Model Name	***MSE***	***MAE***	R2
ARIMA	18.3776	4.2869	0.8905
Xgboost	3.8974	1.9741	0.9768
LSTM	4.231	2.057	0.9322
Ns-Transformer	2.6732	1.6349	0.9743
ours	1.867	1.3663	0.9876

4.3 Analysis of prediction result graph

To further underscore the superior performance of this study in comparison to other models, a comparison between predicted values and true values was conducted and visually represented. **Figure 4** depicts the prediction results of the LSTM model with additional data features. The graphical representation indicates a substantial enhancement compared to using LSTM alone, although the results exhibit a tendency towards relative stability. In **Figure 5**, the visualization of the ns-Transformer model's prediction effects reveals its effec- tiveness in handling non-stationary elements compared to the LSTM model. However, there are still deviations in accuracy observed for certain prediction points. **Figure 6** presents the predicted results of this study, showcasing a notable improve- ment compared to the first two sets of prediction outcomes.

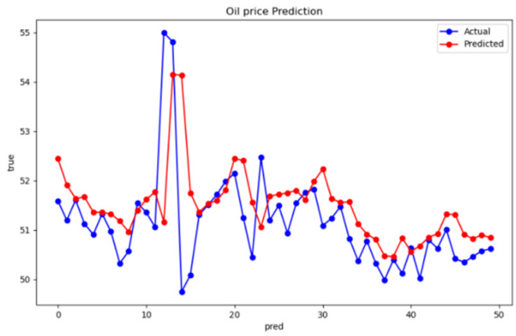

Fig. 4. LSTM prediction effect diagram

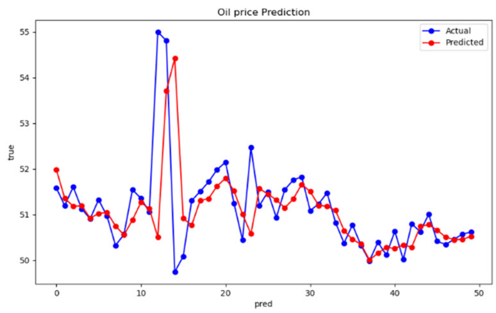

Fig. 5. Ns-transformer prediction effect diagram

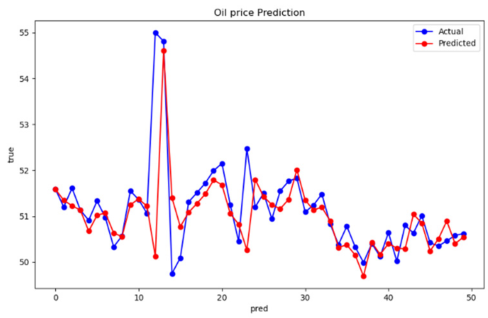

Fig. 6. The prediction effect diagram of this research method

5. Conclusions

This article delves into the challenge of predicting crude oil prices by considering their intrinsic characteristics. Diverging from prior research, it incorporates a more comprehensive set of data features that pertain to the commodity attributes of crude oil prices, while also acknowledging the non-stationary nature inherent in crude oil price data. Employing an effective method to preserve the non-stationary information of the data,the study aims to enhance the predictability of crude oil prices and the overall predictive capability of the model.The experiments conducted demonstrate the efficacy of our method on crude oil price data. In the future, we intend to explore additional relevant technical indicators and employ more effective methods to address data stabilization issues associated with crude oil prices.

References

[1] Dale S. BP statistical review of world energy[J]. BP Plc, London, United Kingdom, 2022: 14-16.

[2] George EP Box and Gwilym M Jenkins. Some recent advances in forecasting and control. Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 17(2):91–109, 1968.

[3] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Mod- eling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research Devel-
opment in Information Retrieval*, pages 95–104. ACM, 2018.

[4] Sepp Hochreiter and Jurgen Schmidhuber.Long short-term memory.*Neural computation*, 9(8):1735–1780, 1997.

[5] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder- decoder approaches. arXiv preprint arXiv:1409.1259*, 2014.

[6] Urvashi Khandelwal, He He, Peng Qi, and Dan Jurafsky. Sharp nearby, fuzzy far away: How neural language models use context. arXiv preprint arXiv:1805.04623*, 2018.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems, pages 5998–6008, 2017.

[8] Liu Y, Wu H, Wang J, et al. Non-stationary transformers: Re- thinking the stationarity in time series forecasting[J]. arXiv preprint arXiv:2205.14415, 2022.Ll

[9] Karasu S, Altan A. Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization[J].
Energy, 2022, 242: 122964.

[10] Lin Lin, Wang Fang, Xie Xiaolong, et al. Random forests-based extreme learning machine ensemble for multi-regime time series prediction[J]. Expert Systems with Applications, 2017, 83(22): 164–176.

[11] Gupta V, Pandey A. Crude Oil Price Prediction Using LSTM Net- works[J]. 2018(3).

[12] Graham Elliott, Thomas J. Rothenberg, and James H. Stock. Efficienttests for an autoregressive unit root. Econometrica.1996.