DMEformer: A newly designed dynamic model ensemble transformer for crude oil futures prediction

Chao Liu [a], Kaiyi Ruan [b], Xinmeng Ma [c,*]

[a] College of Business and Trade, Hunan Industry Polytechnic, Changsha, 410208, China
[b] School of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan, 030024, China
[c] School of Energy, Power and Mechanical Engineering, North China Electric Power University, Beijing, 102299, China

ABSTRACT

Crude oil futures prediction plays an important role in ensuring sustainable energy development. However, the performance of existing models is not satisfactory, which limits its further application. The poor performance mainly results from the lack of data mining of economic models and the poor stability of most data analysis models. To solve the above problems, this paper proposes a new dynamic model ensemble transformer (DMEformer). The model uses three different Transformer variants as base models. It not only ensures the difference of base models but also makes the prediction results of base models not to appear disparity. In addition, NSGA-II is adopted to ensemble the results of base models, which considers both the modeling stability and accuracy in the optimization. Finally, the proposed model adopts a dynamic ensemble scheme, which could timely adjust the weight vector according to the fluctuation of energy futures. It further improves the reliability of the model. Comparative experiments from the perspective of single models and ensemble models are also designed. The following conclusions can be drawn from the experimental results: (1) The proposed dynamic ensemble method can improve the performance of the base model and traditional static ensemble method by 16% and 5% respectively. (2) DMEformer can achieve better performance than 20 other models, and its accuracy and MAPE values were 72.5% and 2.8043%, respectively. (3) The proposed model can accurately predict crude oil futures, which provides effective support for energy regulation and sustainable development.

1. Introduction

1.1. Background

As countries around the world pay more and more attention to environmental protection, carbon emission control is favored by scholars and governments [1]. At present, many countries have put forward different carbon emission control strategies. China has set the goal of "peaking carbon" by 2030 and "carbon neutrality" by 2060 [2]. EU puts forward an energy system ensemble strategy. However, it would be a positive choice starting from the economy to achieve carbon emission control. The government's macro-regulation of the carbon trading market is an effective means to reduce carbon emissions [3].

To grasp the market trends and analyze the trend of energy prices is the premise of accurate regulation. At present, many scholars have put forward a variety of energy price forecasting methods from economic and data analysis perspectives. However, the carbon trading market contains many influencing factors, including carbon production, new energy, national policies, social and economic conditions, and so on [4]. Complex factors increase the difficulty of energy futures price prediction and the accuracy and stability of most energy price prediction methods are not satisfactory.

Energy futures price forecasting has statistical methods based on economic theory and intelligent algorithms based on data analysis [5]. Statistical forecasting methods based on economic theories include the information theory model, Bayesian model, Pooled Mean Group (PMG) panel ARDL estimation method [6], and so on. However, this kind of method cannot fully analyze the information from the data. Intelligent prediction methods based on data analysis mainly include machine learning, deep learning, ensemble learning, and so on [7]. Several models like LSTM [8], MOPSO [9], SRU [10], and so on are proven to be useful. However, there is still room for improvement in terms of the efficiency and stability of the models. The recently proposed Transformer-based models have been proven to be more effective than traditional deep networks in natural language processing and time series forecasting [11]. Thus, combining ensemble strategy, this paper explores a Transformer-based ensemble energy timing forecasting method to provide accurate and reliable price forecasting information for carbon trading control.

1.2. Related literature review and our contribution

Up to now, many scholars have carried out various research on energy prices, hoping to analyze the internal factors of price changes or put forward more accurate price forecasting methods. Most of the methods are statistic models and information-based economic models. Atalla et al. analyzed the relationship between economic policies and fossil fuel prices and adopted the Dynamic Stochastic General Equilibrium model to simulate its relevant changes [12]. However, the method neglected the other factors which also had a great impact on the fuel prices. Drachal et al. adopted various Bayesian models and proposed an information theory method to predict the energy price, which has a better effect than ARIMA [13]. But the model was not able to mine deep features in the data. Mensah et al. adopted the ARDL method to simulate the relationship among fossil carbon dioxide emissions, fuel energy consumption, and oil price. He found that there was a long-term bilateral causal relationship between fossil energy consumption and social and economic growth [6]. Liu and Jin used a standard economic method to analyze the relationship between China's electricity market and carbon trading market and found that electricity price was strongly correlated with coal price, but not significantly correlated with other fossil energy prices [14]. Many economists have taken a keen interest in the underlying dynamics of energy prices. The relationship between carbon trading and macro-economy, electricity, automobile trading, energy conservation, and emission reduction has been studied a lot [15–18], but most models could only be adapted to economic forecasting in a certain period or only pay attention to the part of the correlation information, which is not comprehensive and stable enough.

To fully explore the information contained in the data itself, some scholars tried to use intelligent algorithms to realize the time series prediction of energy prices. Many intelligent algorithms such as traditional machine learning models [19] and deep networks have been evaluated in the field of energy economic prediction [20]. Sadorsky applied traditional machine learning methods to predict the stock price of clean energy and verified that support vector machine (SVM), Random Forest (RF), and other models have better prediction performance compared with Lasso and Bayes [21]. But the mentioned models failed to get satisfying results when it comes to other stock price data. Seçkin and his team proposed a multi-objective feature selection method for oil time series prediction [9], which could reasonably seek out the relevant features but it might also lose information in the feature selection process. Li et al. combined variational mode decomposition and random sparse Bayesian learning and achieved positive performance in crude oil prediction [22]. However, the decomposition results from VMD usually changed a lot when the length of the time series is growing which to some extent would affect the prediction accuracy. Sun and his team worked out a secondary decomposition-reconstruction-ensemble approach to forecast crude oil prices [23]. The experiments demonstrated that the data information had been well extracted and the model was effective. Many other models also have been studied for crude oil price forecasting, which can be found in Ref. [24]. Although the above model can realize the prediction of energy futures prices to some extent, the prediction accuracy is not satisfactory due to the solidification of the model and the defects of the traditional algorithm.

Many newly proposed deep network models and intelligent algorithms have been shown to work well in other fields. Although most of them have not yet been applied to the field of energy price forecasting, many models with excellent performance in temporal forecasting have begun to be noticed by economic scholars. Seçkin and Aytac proposed a special LSTM whose parameters were optimized by the chaotic Henry gas solubility optimization [8]. It was able to give a precise prediction of crude oil time series but LSTM was also restricted for the reason that it cannot model long-term dependences of the series. Sabri designed an ADARNN model for oil price forecast and it proved that the model had a favorable performance in a data-rich environment [25]. But the method was also restrained because of the data resources and richness, neglecting the information mining of price data itself. Teng et al. proposed a hierarchical attention-based LSTM model [26]. The attention mechanism can give different degrees of attention to the data in different periods, and realize more accurate prediction of stock market trends. Sinha uses a three-dimensional convolutional neural network (CNN) to predict the trend of stock plates, but a convolutional neural network lacks the learning of time sequence information [27]. Sadorsky et al. verified the futures price prediction effect of various tree-based models and found that the tree model has a positive effect in dealing with short-term prediction [28]. However, the tree-based models are poor in long-term prediction. Jafari and his team verified the predictive performance of GCN in terms of stock rise and fall and found that the graph structure has advantages in processing diversified information [29]. But the graph design of GCN is controversial. The graph structure designed for the market is not universal and the edge structure may not have an exact correlation. In addition, many other intelligent algorithms have also been tested for their performance in the economic field, such as long-term memory networks (LSTM), hybrid models, etc. [30–33]. But the

performance and universality are still under improvement. Thus, inventing more stable and effective models still appeals to many intelligent algorithm researchers.

Recently, Transformer-based models have been proven to be effective in natural language processing and time series forecasting [11]. It mainly relies on the attention mechanism to model the relationship between the input series and output series. Wang and his team proposed a deep Transformer model for stock market index prediction [34]. But it was unstable enough when there were steep changes in the index data. Kwon and Lee designed a recurrent Transformer model for coil temperature forecasting which had a good performance but also resulted in more computing burden [35]. Many other Transformer-based models have also been studied by researchers like CL-former [36], OD-former [37], and so on, which can adapt to various time series forecasting issues. All the above models have achieved favorable performance in other fields, thus, Transformer-based models are evaluated in the paper to validate the performance in crude oil prices forecasting. Besides, most of the models are single models or application models in specific scenarios, so the stability is poor when it comes to the complex and changeable situation of the energy futures market.

To solve the above problems, some scholars adopt an ensemble model or hybrid model to solve the above problems. Toochaei and Moeini tested multiple ensemble classifiers including Random Forest, LightGBM, XGBoost, Extra-Trees, AdaBoost, and CatBoost, and found that ensemble classifiers were more robust than other models [38]. However, all the mentioned models are ensemble by simple tree models, and the prediction accuracy still has room for improvement. Liu et al. proposed a hybrid ensemble deep reinforcement learning model to solve the problem of timing prediction [39]. He used three different deep network models as the base model to learn time series from different angles, which has good accuracy and robustness. Many other ensemble models or mixed models have also been studied [40–42]. But on the one hand, the base models are generally completely different, so the prediction results of base models may be quite different from each other. On the other hand, most ensemble methods ignore the impact of the new changes in the time series on the model, and cannot adjust the model to adapt to the changes.

Therefore, this paper takes three Transformer variant models as base models and adopts the dynamic ensemble method to learn time series changes, which can ensure the stability and accuracy of energy futures price prediction. **The innovation of this model is mainly reflected in the following three aspects:**

(a) Autoformer is able to decompose the time series, mining inner information, and has a good performance in long-time series prediction. Reformer is an efficient Transformer, which is effective and consumes less computing resources. And Informer can not only study the relationship between longer series but also reduce the accumulated error of the model. Taking the three same-origin Transformer variant models as base models not only ensures the difference among the base models but also prevents the case that the forecast results of the base models are far apart.

(b) Using a multi-objective genetic algorithm (NSGA-II) to learn the weight of base models, taking the deviation and variance of the model in the process of cross-validation as the target, can ensure the generalization performance and prediction accuracy of the model.

(c) The dynamic ensemble method selects the optimal weight of the previous time from the Pareto plane as the weight of base models at the prediction time, which can timely adjust the model according to the fluctuation of energy futures, and further improves the robustness of the model.

The paper is organized as follows: In the first two parts, the significance of energy futures prediction and related research are expounded. Section 3 introduces the proposed dynamic ensemble Transformers model in detail. Section 4 is the experimental part. The paper compares the performance of different advanced models and different ensemble models to verify the effectiveness of the proposed models. Finally, Section 5 summarizes the results of this study and expounds the future research work.
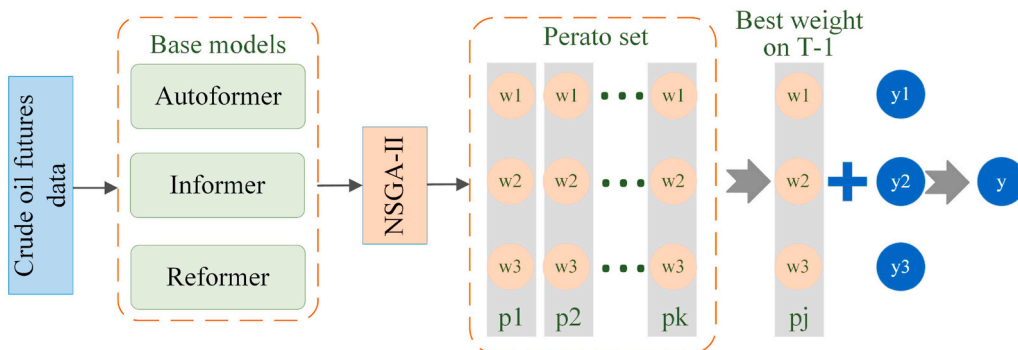

Fig. 1. The framework of the proposed model.

2. Methodology

2.1. Problem analysis and model framework

Energy future price change The prediction of energy price change is a typical time series prediction problem. Different from general time series, energy futures prices have the characteristics of large manual intervention and small randomness. Usually, it is influenced by recent policies or cyclical transactions that produce changes different from the historical sequence. The dynamic ensemble diversified Transformer model proposed in the paper can give different perspectives to historical sequences and mine deep features in the data. In addition, the dynamic ensemble method allows prediction patterns to change dynamically with sequence changes so that the prediction process has a high adaptability to the new changes in energy prices, which further improves the prediction accuracy.

As shown in Fig. 1, the proposed model mainly includes three base models based on Transformer and a more robust dynamic ensemble strategy. The base model includes Autoformer, Reformer, and Informer. Autoformer has remarkable performance in long-time sequence and can focus on long-time sequence information. The reformer greatly improves the training efficiency of the transformer. Informer is more balanced on the short-and-long-term information of time series, and it solves memory and efficiency challenges. By taking the optimal weight of the last moment as the ensemble weight of the next moment, the dynamic ensemble strategy improves the dynamic adaptability of the model and could predict the change of oil futures more accurately.

2.2. Base models

2.2.1. Autoformer

Autoformer is a long-time series prediction model based on deep decomposition architecture and autocorrelation mechanism, which was proposed by Wu and his team [43]. In recent years, Transformer has shown positive performance in the field of time series prediction and attracted the attention of scholars all over the world. However, there is still room for improvement in accuracy and efficiency. On the one hand, with the prolonging of predictive timeliness, it is difficult to find reliable temporal dependence from complex time patterns by direct use of the self-attention mechanism. On the other hand, due to the quadratic complexity of self-attention, the sparse version of the model has to be used, which will limit the efficiency of information utilization and affect the prediction effect [44]. To solve the above problems, deep decomposition architecture and automatic correlation mechanisms are adopted in Autoformer.

As shown in Fig. 2, Autoformer's framework consists of an encoder and decoder, but unlike a typical transformer, the encoder, and decoder are embedded with a decomposition unit and an AutoCorrelation module. The decomposition unit is mainly based on the idea of moving average, which can smooth the time series and separate the period term from the trend term. Since the decomposition process is embedded in the encoding and decoding process, the model gradually separates the trend term and the period term from the hidden variables in the prediction process, which is a gradual decomposition. In addition, the prediction result optimization and sequence decomposition are carried out alternately by the model, which can realize the mutual promotion. The AutoCorrelation module mainly includes Period-based dependencies and Time delay aggregation. It can find similar subprocesses of similar phases of different cycles, so as to achieve efficient sequence-level joining. Besides, it also has the ability to learn the long-dependent correlation in time series.

2.2.2. Reformer

The general transformer model tends to have a large memory footprint when dealing with long sequences. To conduct efficient
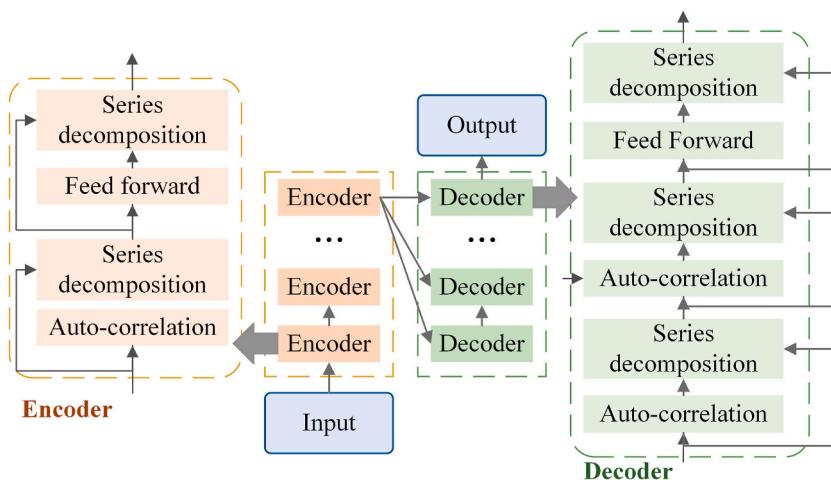

Fig. 2. The structure of the Autoformer part.

training, Reformer has improved Transformer in three parts [45]. Firstly, locality-sensitive hashing attention (LSH Attention) is used to replace the original attention layer, which can improve the parallelization processing speed of the attention layer. Secondly, reversible residuals are proposed to reduce the memory footprint caused by the activation function computation. Finally, it blocks the input and calculates one module at a time to prevent the hidden layer from being too large. Due to the improvement, Reformer needs less computing burden and is obviously more efficient than Transformer. Fig. 3 shows the internal architecture of the Reformer's encoder and decoder.

The Reformer's principle of extreme simplification of training and memory is given as follows:

(1) LSH Attention

When dealing with the time series prediction problem, the common attention layer mainly adopts the form of a shrinking dot product, while the LSH attention layer proposes the method of calculating the points in high dimensional space by the hash function. The hash function is used to divide the input feature vector into two different hash buckets, and the sequence is reordered according to the hash buckets. The new sequence is then partitioned and the attention network is trained in parallel.

(2) Reversible residual network

The residual block of the Transformer consists of an attention layer and a feedforward layer. Therefore, each layer activation function needs to save the input to calculate the backpropagation gradient, which takes up a large amount of storage space. The reversible residual network can realize the reverse recalculation of the input of the activation function so that the activation function can be calculated only once in the first layer.

(3) Block forward network

Block the input to reduce the amount of computation per training, thus reducing memory consumption. And they can be obtained by Eq (1) and Eq (2).

$$Y_2 = \left[Y_2^{(1)}, Y_2^{(2)}, \cdots, Y_2^{(M)} \right] \tag{1}$$

$$Y_2^{(j)} = X_2^{(j)} + FFN\left(Y_1^{(j)} \right) \tag{2}$$

where Y represents the output of the activation function, X represents the input of the activation function, and FFN represents the fully connected layer. The superscript (j) represents the j-th block.

2.2.3. Informer

Informer was proposed by Zhou and his team to improve Transformer from the perspective of solving the computational complexity and input characteristics of long time series [46]. On the one hand, ProbSparse self-attention was chosen, which reduced the computational complexity and spatial complexity of conventional self-attention. Besides, self-attention distilling was put forward to compress the model. By shortening the input sequence length of each layer, the memory usage of J stacked layers can be reduced, and the effect of pruning can be achieved. What's more, this model adds the processing of date, taking the year, month, and day time of date as features, and improves the decoder of the generation, which can output the prediction results in one step. The specific principles are as follows:
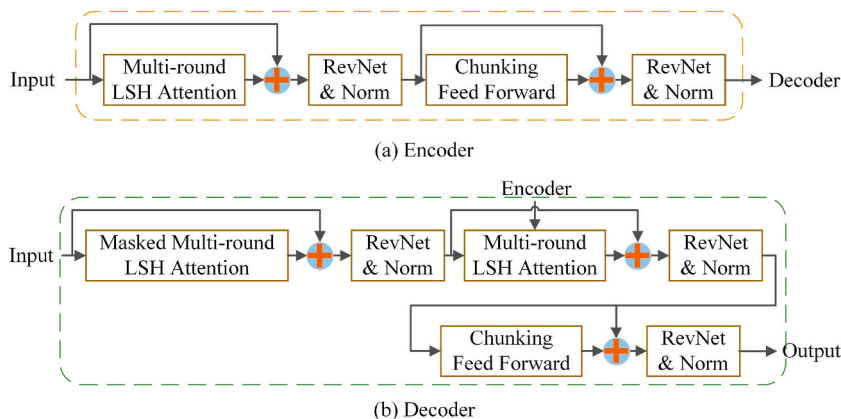

Fig. 3. The encoder and decoder of Reformer.

(1) ProbSparse self-attention

The main idea of this method is to sample 5*ln$L$ points for dot product calculation. Partial keys were randomly sampled for each query, and the sparsity score M of each query was calculated and sorted. The 5*ln$L$ queries with the highest score are selected to calculate the dot product, and the average value of the input of the other queries in this layer is taken as the output. Through the above methods, the algorithm complexity can be reduced to $O(L\ln L)$, which greatly improves the algorithm running efficiency. But it is worth noting that this method is only applied to the self-attention layer of the encoder.

(2) Self-attention Distilling

The main principle of this mechanism is to add a 1-dimensional convolution layer and maximum pooling layer after each attention block, which can reduce the feature dimension. At the same time, each sample sequence has its input length reduced by half at input time by 1-dimensional convolution. The above distillation operation further reduces the algorithm complexity.

(3) Generative decoder

Informer's decoder is different from a general Transformer that a generative decoder can generate all the predictive output at once. The input of this decoder is a concatenated sequence of the tail of the encoder input and the 0 matrices of the same dimension as the target to be predicted, and the tail of the output is the forecast result.

Autoformer, Reformer, and Informer have their characteristics for long-time series prediction and can analyze and learn the series from different perspectives. By integrating the three base models, their respective advantages can be combined to ensure stability and improve the accuracy of the oil futures forecast.

2.3. The dynamic ensemble strategy

The dynamic ensemble strategy mainly consists of two parts: one is to obtain the optimal weight Pareto surface through multi-objective optimization; the other is to realize dynamic weight update by selecting the weight vector corresponding to the optimal result at the last moment on the Pareto surface. The multi-objective dynamic ensemble method can not only improve the stability and accuracy of the model for long sequence prediction but also adapt to the new sequence changes in time.

2.3.1. NSGA-II

NSGA-II is a multi-objective optimization algorithm based on a genetic algorithm, proposed by Deb [47]. It is widely used in complex and nonlinear multi-objective optimization problems and has the advantages of uniform solution distribution, many diversity, and elite strategy. NSGA-II can find the Pareto solution set of the multi-objective optimization problem and obtain the Pareto optimal solution. Compared with the NSGA algorithm, this algorithm uses fast non-dominated sorting, which has a faster running speed and a lower complexity. The specific algorithm principle is clarified as below:

Step 1. Generate the first-generation parent population Q_0 with a size of N using random numbers.
Step 2. Perform fast non-dominated sorting for the parent population. The process of fast non-dominated sorting is as follows: (1) Set two parameters, $n(i)$ and $S(i)$, where the former is the number of individuals dominating individual i in the population and the latter is the set of individuals dominated by individual i. (2) All the individuals satisfying the condition of $n(i) = 0$ in the population are stored in the set $O(1)$, and the rank of all the individuals in $O(1)$ is set as 1. (3) Traverse the set $S(j)$ of all individuals dominated by individual j in $O(1)$, and subtract 1 from $n(k)$ of each k in $S(j)$. If $n(k)$-1 = 0, then the individual k is placed in the set $O(2)$, and all the individuals in $O(2)$ rank 2. (4) Operation (3) is performed on set $O(2)$ until the rank of all individuals has been determined.
Step 3. Select, cross, and mutate Q_0 to generate the first generation of offspring population.
Step 4. Mix the parent and child populations to produce a new population D_0 with a size of 2 N.
Step 5. Calculate the crowding degree of D_0 and conduct fast non-dominated sorting. Select N individuals with the best performance to form a new parent group Q_1.
Step 6. If the end condition is met, terminate the cycle and output the optimal solution set; otherwise, go to Step 4.

2.3.2. Dynamic-NSGA2 strategy

(1) Multi-objective function

The reasonable objective function is the key to ensuring the accuracy of the integration model. In this paper, the deviation and variance of the 10-fold cross-validation results of the integrated model are taken as the objective function.

Model deviation: Model deviation can reflect the prediction accuracy of the model, ensuring that the integrated optimization can select the weight vector to make the results more accurate. The deviation can be obtained by Eq (3).

$$\min f_1 = \frac{1}{n_v} \sum_{i=1}^{n_v} \left(Y_i - \frac{1}{k} \sum_{j=1}^{k} \widehat{Y}_{ij} \right) \tag{3}$$

where Y_i represents the true value of the i-th sample in the verification set, $\widehat{Y}_{ij}$ represents the predicted value of the trained model in the i-th sample in the verification set by the j-th fold cross-verification, n_v represents the number of samples in the verification set, and k represents the number of folds in the cross-verification ($k = 10$).

Model variance: Model variance represents the fluctuation of the predicted value of 10-fold cross-validation, which reflects the stability of the model. By designing this function, integrated optimization tends to select the model with low fluctuation of prediction results. And it can be obtained by Eq (4).

$$\min f_2 = \frac{1}{n_v} \sum_{i=1}^{n_v} \left(\frac{1}{k} \sum_{j=1}^{k} \left(\widehat{Y}_{ij} - \frac{1}{k} \sum_{j=1}^{k} \widehat{Y}_{ij} \right)^2 \right) \tag{4}$$

(2) Dynamic integration method

Algorithm 1. The test process of crude oil prices prediction at t moment

Input: t-5 to t-1 series of crude oil prices: X_t

t-6 to t-2 series of crude oil prices: X_{t-1}

The Perato weight set $P = \{p_1, p_2, \cdots, p_l\}, p_l = (w_1, w_2, w_3)$ obtained from NSGA-II on the validate set

Output:

Final forecasting results at t moment: $\widehat{Y}_t$

Algorithm:

01: Autoformer $\leftarrow X_t, X_{t-1}$

02: Forecasting results from Autoformer at t and t-1 moment $\widehat{Y}_{At}, \widehat{Y}_{At-1}$.

03: Reformer $\leftarrow X_t, X_{t-1}$

04: Forecasting results from Reformer at t and t-1 moment $\widehat{Y}_{Rt}, \widehat{Y}_{Rt-1}$.

05: Informer $\leftarrow X_t, X_{t-1}$

06: Forecasting results from Informer at t and t-1 moment $\widehat{Y}_{It}, \widehat{Y}_{It-1}$.

07: **if** t is the initial moment **do**

08: Randomly select p from P as the weight vector at t moment: $p_t = \widehat{p}$

09: **else do**

10: Calculate $\widehat{Y}_{t-1}$ on each weight vector p in P

$\widehat{Y}_{t-1} = w_1 * \widehat{Y}_{At-1} + w_2 * \widehat{Y}_{Rt-1} + w_3 * \widehat{Y}_{It-1}$

11: Compute the error value

$Error = |Y_{t-1} - \widehat{Y}_{t-1}|$

12: Select out the optimal p with the minimal error at t-1 moment, called $p_{optimal\ t-1}$. $p_t = p_{optimal\ t-1}$

13: **end if**

14: Compute the final prediction results at t moment using p_t.

$\widehat{Y}_t = (\widehat{Y}_{At}, \widehat{Y}_{Rt}, \widehat{Y}_{It}) \bullet p_t^T$

As shown in Algorithm 1, dynamic integration aims to adjust the weight of the model slightly to make the model more adaptable to the prediction under different conditions. Let P represent the Pareto solution set in the training process of the integrated model, and p represents the elements in the solution set, namely the weight vector, then the weight vector selection process of dynamic integration can be obtained by Eq (5).

$$p_t = \begin{cases} \widehat{p}, t = 0 \\ p_{optimal\ t-1}, t > 0 \end{cases} \tag{5}$$

where t is the order of prediction point, p_t is the weight vector for t moment, $p_{optimal\ t-1}$ is the best weight vector for t-1 moment which has the minimal prediction error at t-1 moment, $\widehat{p}$ represents randomly selecting a weight vector from P.

By performing the proposed dynamic ensemble strategy, a stable and accurate crude oil price prediction can be obtained. Besides, all of the important model parameters have been given in Appendix A.

3. Case study

3.1. Crude oil futures dataset

Selecting appropriate data sets is the key to evaluating the performance of the model. The data set adopted in this paper is the historical data of WTI crude oil futures of the US contract for difference. The dataset source URL is shown below: (https://cn.investing.com/commodities/energy). Data from January 1, 2019, to October 3, 2022, were used to conduct modeling and simulation studies.

The data set has a total of 1000 sample points. And the time interval between sample points is daily. The fluctuation characteristics of the dataset are shown in Fig. 4. The training set, validation set, and test set are divided according to the ratio of 3:1:1. The modeling and simulation platform of this paper is Python 3.9.11. In addition, PyTorch 1.11 is used to implement neural network modeling. In this paper, the z-score standardized method was used to preprocess the original data. And all hyperparameters of different models were obtained by the ten-fold cross-validation and raster search.

3.2. Performance evaluation indexes

The time series error evaluation index can effectively analyze the effect of different crude oil futures forecasting models. Three classic indexes, including the MAE (Mean Absolute Error), the MAPE (Mean Absolute Percentage Error), and the RMSE (Root Mean Square Error), are used to evaluate the performance of the proposed dynamic ensemble learning model and all baselines [48]. These time series error evaluation indexes are defined in Eq (6).

$$\begin{cases} MAE = \left(\sum_{T=1}^{n} \left| Y(T) - \widehat{Y}(T) \right| \right) / n \\ MAPE = \left(\sum_{T=1}^{n} \left| \left(Y(T) - \widehat{Y}(T) \right) \middle/ Y(T) \right| \right) / n \\ RMSE = \sqrt{\left(\sum_{T=1}^{n} \left[Y(T) - \widehat{Y}(T) \right]^{2} \right) / n} \end{cases} \tag{6}$$

where $Y(T)$ represents actual crude oil futures data. $\widehat{Y}(T)$ represents the crude oil futures data got by the ensemble prediction model. n represents the number of samples.

In addition to the above evaluation indexes, the directional forecasting (accuracy of prices going up or down) is also an important index. This index can assess the model's ability to correctly judge price trends. And it can be obtained by Eq (7).

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

where TP stands for a true positive example. TN stands for true negative example. FP refers to false positive examples. FN refers to false negative examples.

At the same time, it is important to visually analyze the performance differences between different models. In order to effectively compare the performance of different algorithms, the Promoting percentages of the MAE (P_{MAE}), the Promoting percentages of the MAPE (P_{MAPE}), and the Promoting percentages of the RMSE (P_{RMSE}) are used [49]. These indexes can be obtained by Eq (8).
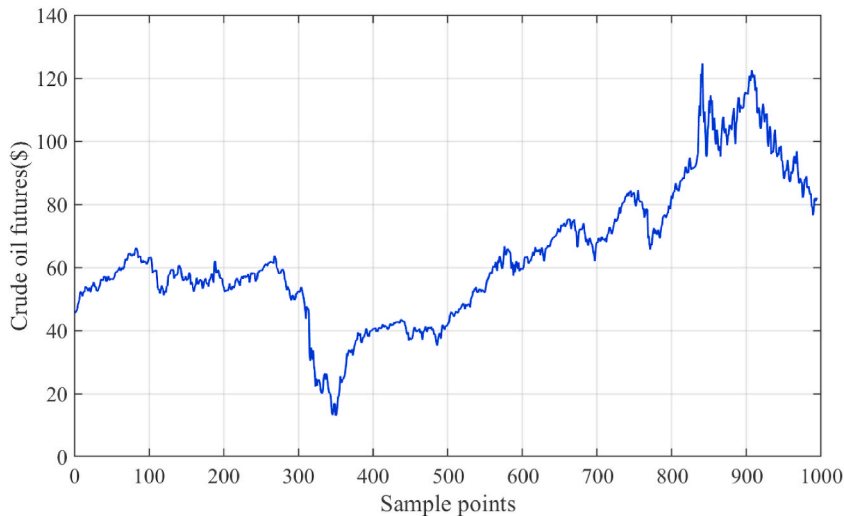

Fig. 4. The raw crude oil futures time series.

$$
\begin{cases}
P_{MAE} = \dfrac{(MAE_a - MAE_b)}{MAE_a} \\[2mm]
P_{MAPE} = \dfrac{(MAPE_a - MAPE_b)}{MAPE_a} \\[2mm]
P_{RMSE} = \dfrac{(RMSE_a - RMSE_b)}{RMSE_a}
\end{cases}
\tag{8}
$$

3.3. Experimental results and comparative analysis with benchmark algorithms

3.3.1. Comparative experimental results of different predictors

To fully compare and select the base predictors with the best performance, this section compares nine existing base predictors, including Transformer based frameworks (Autoformer, Reformer, and Informer), deep learning-based frameworks (TCN, GRU, and LSTM) and traditional neural network frameworks (MLP, ELM, and RBF). Table 2 fully illustrates the prediction errors of these base models. Fig. 5 shows the prediction results of the three models based on the Transformer. From Table 1 and Fig. 5, the following conclusions can be drawn:

(1) Compared with Transformer and deep learning models, the prediction error of traditional neural networks is larger. This proves the effectiveness of deep learning and Transformer in oil futures time series modeling. The possible reason is that the structure of multiple hidden layers effectively improves the ability of neural networks to mine deep hidden information.
(2) Compared with the transformer-based models and the deep learning models, the transformer-based models can achieve better experimental results. This fully proves the effectiveness of force transformers in the field of time series prediction. The feasible reason is that the attention mechanism and time series encoding structure effectively improve the ability to extract original temporal features, which ensures the ability of predictive modeling.
(3) Different types of transformer-based prediction models have different performances. This fully proves that different predictors have different adaptability to complex and changeable time series. To fully improve the overall ability of the model to adapt to the data, ensemble learning was adopted to improve the performance.

3.3.2. Comparative experimental results of different ensemble methods

In this section, to fully prove that the dynamic ensemble learning method proposed in this paper has excellent weight decision performance, this paper constructs comparative experiments from the following three perspectives:

Part I: By comparing the prediction results of ensemble learning and three Transformer based models, it can fully prove that the ensemble learning method can effectively improve the comprehensive strength of the model.
Part II: By comparing NSGA2 with the traditional optimization algorithm, the adequacy and effectiveness of this method to achieve model weight ensemble are fully proved.
Part III: The ablation experiments show that dynamic weight selection can effectively improve the effect of NSGA2 in integrating different base models on time series forecasting.
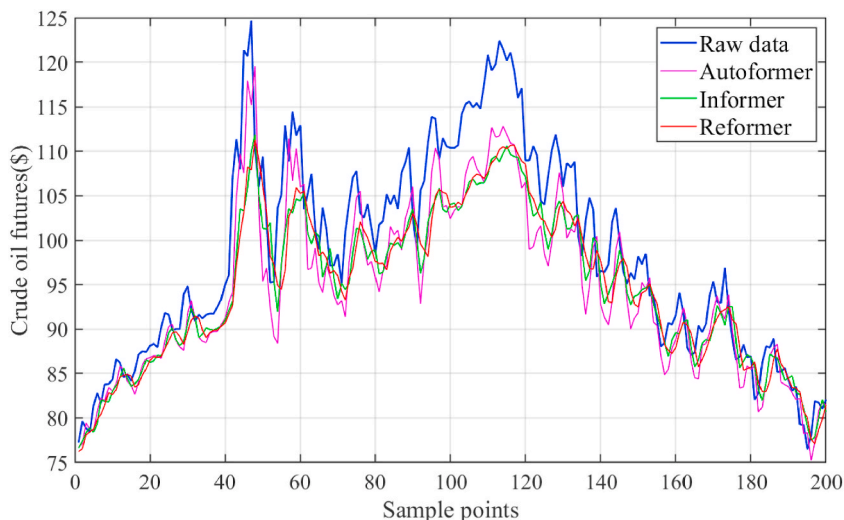

Fig. 5. The prediction results of multiple models.

Table 1
The error evaluation index of different predictors.

Forecasting models	MAE ($)	MAPE (%)	RMSE ($)	ACC (%)
Autoformer	1.0714	3.3863	1.4563	63.5
Reformer	1.1251	3.4590	1.4876	64.0
Informer	1.0678	3.4242	1.5107	63.0
TCN	1.3731	4.1863	1.8016	61.5
GRU	1.3680	4.1908	1.7946	60.5
LSTM	1.3978	4.3200	1.7628	61.0
MLP	1.4535	4.4242	1.9006	58.5
ELM	1.4893	4.6356	2.0031	59.0
RBF	1.5747	4.7860	2.0528	57.5

Table 2
The error evaluation index of different ensemble learning models.

Forecasting models	MAE ($)	MAPE (%)	RMSE ($)	ACC (%)
Dynamic-NSGA2	0.8879	2.8043	1.1588	72.5
NSGA2	0.9522	2.9648	1.2348	69.0
MODE	0.9524	3.0005	1.2353	67.5
BA	0.9658	3.0049	1.2495	66.0
PSO	0.9741	3.0333	1.2825	67.0
GA	0.9819	3.0535	1.2833	66.5

From Tables 2–4 and Fig. 6, the following conclusions can be drawn:

(1) Compared with a single predictor, ensemble learning effectively improves the prediction performance of the model. This fully demonstrates the powerful predictive ability of ensemble learning. Ensemble learning determines the weight by fully analyzing the correlation between the prediction results of the base model and the target, which ensures the accuracy of the prediction.

(2) Compared with other ensemble learning methods, NSGA2 can achieve the best ensemble results. On the one hand, this method uses the idea of the multi-objective ensemble to avoid overfitting. On the other hand, the principle of iterative population search is adopted to ensure the comprehensiveness of the weight search.

(3) Compared with static ensemble learning, dynamic ensemble learning can effectively improve the overall performance of the prediction model. Experimental results demonstrate the effectiveness of dynamic ensemble learning. Based on the greedy principle, the model can dynamically select the Pareto solution set according to the changes in the time series, which improves the ability of the model to adapt to the crude oil futures data.

3.4. Comparing analysis with baselines

To prove that the dynamic ensemble learning model adopted in this paper has excellent ensemble analysis and modeling ability of crude oil futures, this section compares the performance of four baselines. These six baselines include three advanced time series forecasting frameworks (Sun's model [23], Su's model [50], and Verma's model [51]) and three classical forecasting models (LSTNet, SVM, and ARIMA). Figs. 7–10 give the MAE, MAPE, RMSE and Accuracy values of all baselines and the proposed dynamic ensemble learning model. Fig. 11 shows the predicted results of the proposed dynamic ensemble learning model and other baseline models. Table 5 shows the training time of different models. Based on Figs. 7–11 and Table 5, the following conclusions can be drawn:

(1) Compared with the classical prediction methods, the SOTA method proposed by the researchers has better prediction performance. This fully proves that ensemble learning and deep learning can effectively optimize the performance of traditional predictors, which fully ensures the application value of the model in the field of crude oil prediction.

Table 3
The promoting percentages of the ensemble model by single models.

Method	Indexes	Results
Dynamic-NSGA2 vs. Autoformer	P_{MAE} (%)	17.1271
	P_{MAPE} (%)	17.1869
	P_{RMSE} (%)	20.4285
Dynamic-NSGA2 vs.Reformer	P_{MAE} (%)	21.0826
	P_{MAPE} (%)	18.9274
	P_{RMSE} (%)	22.1027
Dynamic-NSGA2 vs. Informer	P_{MAE} (%)	16.8477
	P_{MAPE} (%)	18.1034
	P_{RMSE} (%)	23.2938

Table 4
The promoting percentages of the proposed model by other ensemble models.

| Method	Indexes	Results
Dynamic-NSGA2 vs. NSGA2	P_{MAE} (%)	6.7528
	P_{MAPE} (%)	5.4135
	P_{RMSE} (%)	6.1548
Dynamic-NSGA2 vs. MODE	P_{MAE} (%)	6.7724
	P_{MAPE} (%)	6.5389
	P_{RMSE} (%)	6.1928
Dynamic-NSGA2 vs. BA	P_{MAE} (%)	8.0659
	P_{MAPE} (%)	6.6758
	P_{RMSE} (%)	7.2589
Dynamic-NSGA2 vs. PSO	P_{MAE} (%)	8.8492
	P_{MAPE} (%)	7.5495
	P_{RMSE} (%)	9.6452
Dynamic-NSGA2 vs. GA	P_{MAE} (%)	9.5733
	P_{MAPE} (%)	8.1611
	P_{RMSE} (%)	9.7016
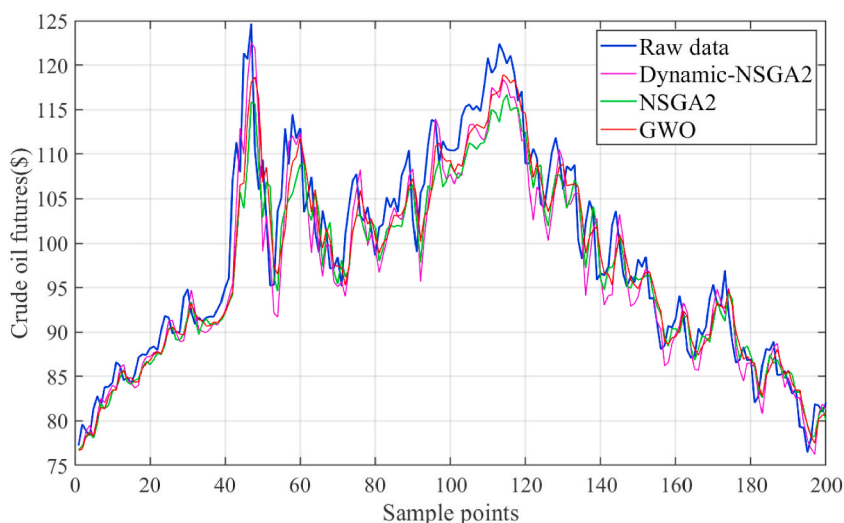
Fig. 6. The prediction results of different ensemble models.
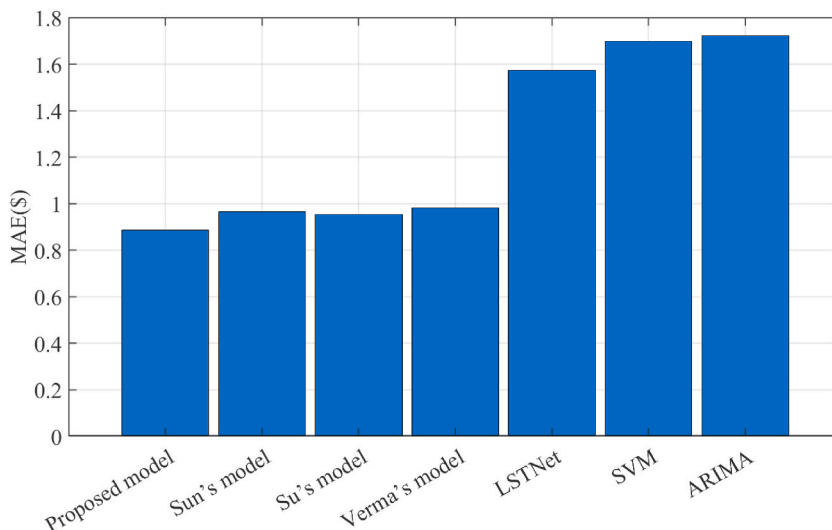
Fig. 7. The MAE values of all baseline and the proposed dynamic ensemble learning model.
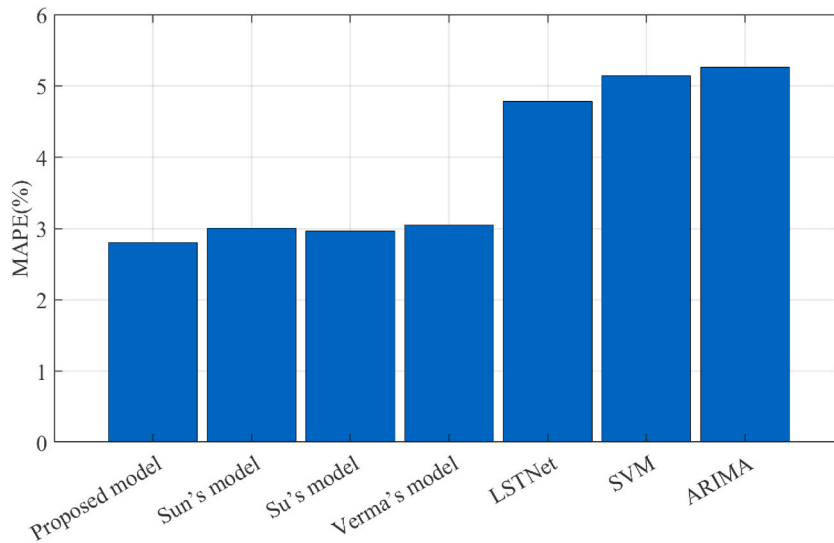
Fig. 8. The MAPE values of all baseline and the proposed dynamic ensemble learning model.
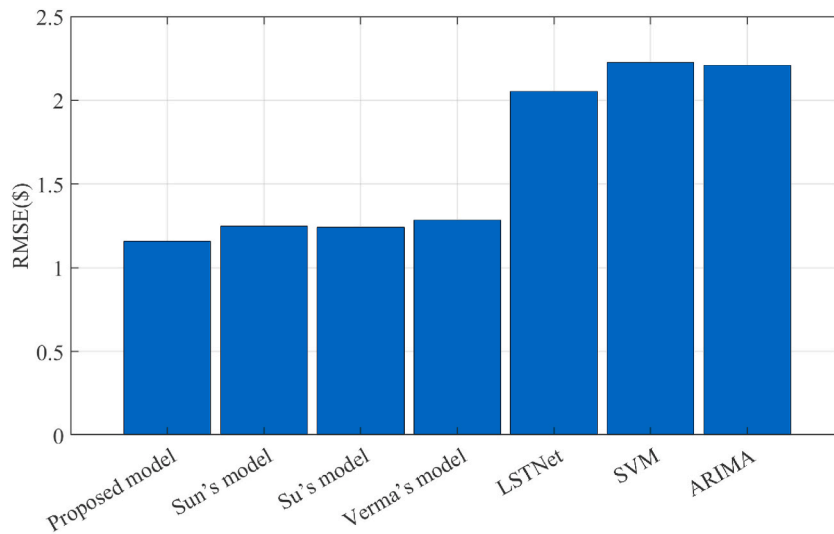
Fig. 9. The RMSE values of all baseline and the proposed dynamic ensemble learning model.

(2) The oil prediction framework based on Transformer and dynamic ensemble learning proposed in this paper can achieve the most satisfactory results. On the one hand, three predictors based on transformers with different frameworks and features can fully analyze the characteristics of oil time series and establish a high-precision forecasting model. On the other hand, the dynamic ensemble learning method combining NSGA2 and the greedy principle can fully analyze the correlation between the three transformers with the real value and realize effective dynamic weight decisions. Therefore, the dynamic ensemble transformer prediction framework proposed in this paper can achieve excellent application prospects in the field of crude oil futures prediction.

3.5. Discussion and analysis

Based on all the experimental results, the practicability and effectiveness of TFEformer in the field of crude oil futures prediction can be proved. All the results are discussed and analyzed in this section.

(1) Compared with traditional CNN, RNN, and shallow neural network methods, the crude oil futures prediction model based on the Transformer can achieve better results. On the one hand, Transformer uses a self-attention mechanism to fully analyze the context dependencies of the time series, which enables the model to effectively analyze the overall information of the series. On
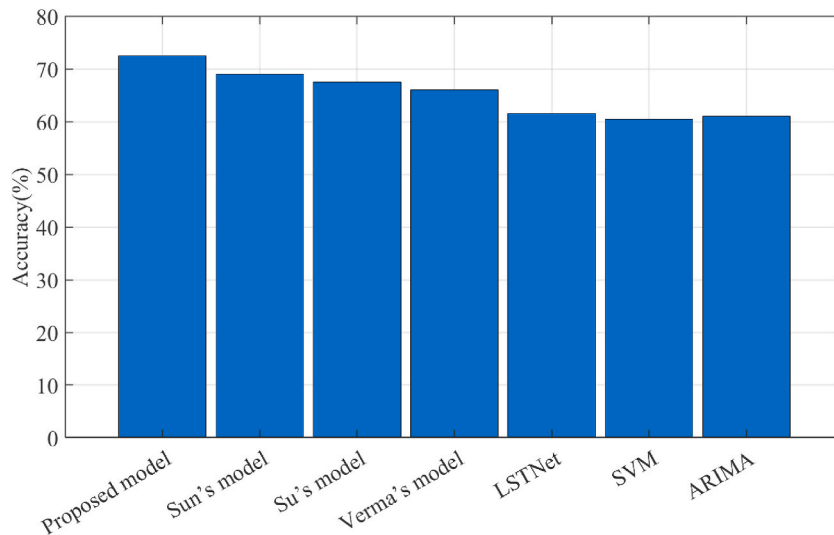
Fig. 10. The Accuracy values of all baseline and the proposed dynamic ensemble learning model.
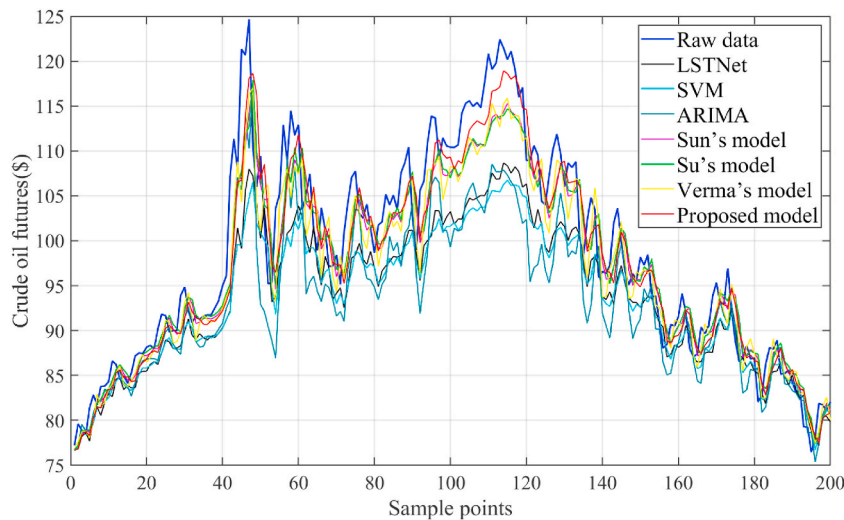

Fig. 11. The prediction results of the proposed model and different baseline models.

Table 5
The training time of different models.

Models	Time (s)
DMEformer	187.51
Sun's model	235.87
Su's model	203.18
Verma's model	52.75
LSTNet	20.18
SVM	10.27
ARIMA	15.45

the other hand, the structure of residual connection and multi-layer depth network ensures the ability of the model to mine the hidden information in the original data, which further improves the modeling effect.

(2) Compared with a single model, the ensemble learning model can fully improve the generalization of the model and further improve the prediction performance of the model. On the one hand, due to the differences in network structure, different types of Transformer variants have different adaptability to time series, which makes it arduous for a single model to fit all cases. On

the other hand, ensemble learning combines the advantages of different base models to make the ensemble model adapt to more modeling scenarios. Therefore, the ensemble learning framework proposed in this paper can achieve better results than all the base models.

(3) Compared with the static ensemble learning method, the dynamic ensemble learning method can achieve better prediction accuracy. The main reason is that the dynamic ensemble learning framework can dynamically optimize the weights of different base models according to the changes in time series, which makes the models better adapt to the changing characteristics of time series. Therefore, the dynamic ensemble learning framework can effectively improve the overall prediction effect of the model.

(4) Compared with other existing models, the DMEformer proposed in this paper effectively combines the advantages of Transformers and dynamic ensemble learning. Specifically, based on the dynamic ensemble learning strategy, DMEformer fully integrates multiple Transformer variants with different characteristics and fully improves the ability to adapt to the complex and changeable time series of crude oil futures. Therefore, the model can significantly improve the prediction accuracy.

4. Conclusion and future work

The high-precision prediction method of crude oil futures provides effective technical support to ensure the sustainable development of economic energy. To accurately predict the future trend of crude oil futures, this paper proposes a dynamic ensemble Transformer forecasting framework. The core contribution and summary of this paper mainly include the following elements:

(1) In this paper, three kinds of transformers with their characteristics are used as the main base models. Different from the traditional algorithms, the transformer utilizes an attention mechanism, multi-layer encoder, and decoder to fully exploit the depth features of the time series. Therefore, the time series prediction model based on the transformer can achieve satisfactory results.

(2) Combining NSGA2 and the greedy principle, this paper proposes a new dynamic ensemble learning framework for time series prediction. Different from the traditional static ensemble learning, the model can dynamically select the most appropriate solution set from the Pareto solution set through the greedy principle and the changes of the crude oil futures series to ensure the overall modeling effect.

(3) The prediction framework of the dynamic ensemble transformer proposed in this paper can achieve better prediction results than all alternative models and Baselines. All the experiments fully prove the accuracy of the framework in crude oil futures prediction.

The crude oil futures prediction technology plays an important role in ensuring the sustainable development of the economy and energy. In the future, this paper will conduct further research from the following perspectives:

(1) The effective combination of forecasting results and economic policies can provide effective assistance in decision-making and energy development. In the future, it is extremely important to combine economic policy and forecasting models effectively.

(2) The construction of a complete set of big data analysis systems for crude oil futures is of great significance for realizing the implementation of the model and helping decision-makers to analyze the data. In the future, a complete set of energy data analysis and economic sustainability analysis systems can be implemented by combining big data analysis platforms.

Funding

Author contribution statement

Chao Liu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Kaiyi Ruan: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Xinmeng Ma: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

Data associated with this study has been deposited at https://cn.investing.com/commodities/energy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Model hyperparameters

The hyperparameters of different models are given in Table. A.

Table. A
The hyperparameters of different algorithms.

Name of parameter	Selected parameter
NSGA2	
Maximum iteration	100
Number of search agents	50
Number of objective functions	2
Exponential distribution of crossover and mutation	20
MODE	
Maximum iteration	100
Number of search agents	30
Number of objective functions	2
Crossover probability f	0.5
PSO	
Maximum iteration	100
Inertia weight	0.8
Accelerated constant	2
BA	
Maximum iteration	100
Impluse	0.4
Maximum frequency	0.5
Minimum frequency	0.1
GA	
Maximum iteration	100
Crossover rate	0.3
Mutation rate	0.1
Informer	
Learning rate	0.001
Optimizer	Adam
History length	5
Size of hidden units	256
Training Epochs	100
Autoformer	
Learning rate	0.001
Optimizer	Adam
History length	5
Size of hidden units	128
Training Epochs	100
Reformer	
Learning rate	0.001
Optimizer	Adam
History length	5
Size of hidden units	512
Training Epochs	100
LSTM	
Learning rate	0.001
Optimizer	Adam
History length	5
Number of hidden layers	2
Training Epochs	200
Number of hidden layer units	64, 32
Number of output layer units	1
GRU	
Learning rate	0.001
Optimizer	Adam
History length	5
Number of hidden layers	2
Training Epochs	200
Number of hidden layer units	128, 64
Number of output layer units	1
TCN	
Learning rate	0.001
History length	3
Optimizer	Adam
Filter size	2

Table. A (*continued*)

Name of parameter	Selected parameter
Dropout	0.05
Training Epochs	100
Number of hidden layer units	64
Number of output layer units	1
ELM	
History length	5
Size of hidden units	16
Transfer Function	Sigmoidal function
Size of output units	1
MLP	
History length	5
Training Epochs	200
Size of hidden units	32
Learning rate	0.001
Size of output units	1

References

[1] J. Chen, R. Adhikari, E. Wilson, J. Robertson, A. Fontanini, B. Polly, O. Olawale, Stochastic simulation of occupant-driven energy use in a bottom-up residential building stock model, Appl. Energy 325 (2022), 119890.

[2] H. Liu, C. Yu, H. Wu, C. Chen, Z. Wang, An improved non-intrusive load disaggregation algorithm and its application, Sustain. Cities Soc. 53 (2020), 101918.

[3] Y. Min, Z. Shuzhen, L. Wuwei, Carbon price prediction based on multi-factor MEEMD-LSTM model, Heliyon 8 (12) (2022), e12562.

[4] Measuring Economies, The trouble with GDP, in: The Economist, Briefing, 2016, pp. 21–24.

[5] P. Owusu Junior, A.M. Adam, E. Asafo-Adjei, E. Boateng, Z. Hamidu, E. Awotwe, Time-frequency domain analysis of investor fear and expectations in stock markets of BRIC economies, Heliyon 7 (10) (2021), e08211.

[6] I.A. Mensah, M. Sun, C. Gao, A.Y. Omari-Sasu, D. Zhu, B.C. Ampimah, A. Quarcoo, Analysis on the nexus of economic growth, fossil fuel energy consumption, CO2 emissions and oil price in Africa based on a PMG panel ARDL approach, J. Clean. Prod. 228 (2019) 161–174.

[7] J. Shah, D. Vaidya, M. Shah, A comprehensive review on multiple hybrid deep learning approaches for stock prediction, Intellig. Sys. Applications 16 (2022), 200111.

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

[9] S. Karasu, A. Altan, S. Bekiros, W. Ahmad, A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series, Energy 212 (2020), 118750.

[10] H. Liu, C. Yu, C. Yu, A new hybrid model based on secondary decomposition, reinforcement learning and SRU network for wind turbine gearbox oil temperature forecasting, Measurement 178 (2021), 109347.

[11] M. Castangia, L.M.M. Grajales, A. Aliberti, C. Rossi, A. Macii, E. Macii, E. Patti, Transformer neural networks for interpretable flood forecasting, Environ. Model. Software 160 (2023), 105581.

[12] T. Atalla, J. Blazquez, L.C. Hunt, B. Manzano, Prices versus policy: an analysis of the drivers of the primary fossil fuel mix, Energy Pol. 106 (2017) 536–546.

[13] K. Drachal, Comparison between Bayesian and information-theoretic model averaging: fossil fuels prices example, Energy Econ. 74 (2018) 208–251.

[14] X. Liu, Z. Jin, An analysis of the interactions between electricity, fossil fuel and carbon market prices in Guangdong, China, Energy Sustain. Develop. 55 (2020) 82–94.

[15] M.G. Flammini, G. Prettico, A. Mazza, G. Chicco, Reducing fossil fuel-based generation: impact on wholesale electricity market prices in the North-Italy bidding zone, Elec. Power Syst. Res. 194 (2021), 107095.

[16] L.V. Smith, N. Tarui, T. Yamagata, Assessing the impact of COVID-19 on global fossil fuel consumption and CO2 emissions, Energy Econ. 97 (2021), 105170.

[17] D. Vicknair, M. Tansey, T.E. O'Brien, Measuring fossil fuel reserves: a simulation and review of the U.S. Securities and Exchange Commission approach, Resour. Pol. 79 (2022), 103028.

[18] S. Qiao, Y.J. Dang, Z.Y. Ren, K.Q. Zhang, The dynamic spillovers among carbon, fossil energy and electricity markets based on a TVP-VAR-SV method, Energy 266 (2023), 126344.

[19] H. Liu, C. Chen, Spatial air quality index prediction model based on decomposition, adaptive boosting, and three-stage feature selection: a case study in China, J. Clean. Prod. 265 (2020), 121777.

[20] D.B. Vukovic, S. Ingenito, M. Maiti, Time series momentum: evidence from the European equity market, Heliyon 9 (1) (2023), e12989.

[21] P. Sadorsky, Using machine learning to predict clean energy stock prices: how important are market volatility and economic policy uncertainty? J. Clim. Finan. (2022), 100002.

[22] T. Li, Z. Qian, W. Deng, D. Zhang, H. Lu, S. Wang, Forecasting crude oil prices based on variational mode decomposition and random sparse Bayesian learning, Appl. Soft Comput. 113 (2021), 108032.

[23] J. Sun, P. Zhao, S. Sun, A new secondary decomposition-reconstruction-ensemble approach for crude oil price forecasting, Resour. Pol. 77 (2022), 102762.

[24] R. Kertlly de Medeiros, C. da Nóbrega Besarria, D. Pitta de Jesus, V. Phillipe de Albuquerquemello, Forecasting oil prices: new approaches, Energy 238 (2022), 121968.

[25] S. Boubaker, Z. Liu, Y. Zhang, Forecasting oil commodity spot price in a data-rich environment, Ann. Oper. Res. (2022) 1–18.

[26] X. Teng, X. Zhang, Z. Luo, Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction, Neurocomputing 505 (2022) 92–100.

[27] S. Sinha, S. Mishra, V. Mishra, T. Ahmed, Sector influence aware stock trend prediction using 3D convolutional neural network, J. King Saud Univer. - Com. Inform. Sci. 34 (4) (2022) 1511–1522.

[28] P. Sadorsky, Forecasting solar stock prices using tree-based machine learning classification: how important are silver prices? N. Am. J. Econ. Finance 61 (2022), 101705.

[29] A. Jafari, S. Haratizadeh, GCNET: graph-based prediction of stock price movement using graph convolutional network, Eng. Appl. Artif. Intell. 116 (2022), 105452.

[30] D. Kadrić, A. Aganovic, E. Kadrić, B. Delalić-Gurda, S. Jackson, Applying the response surface methodology to predict the energy retrofit performance of the TABULA residential building stock, J. Build. Eng. 61 (2022), 105307.

[31] W.-C. Lin, C.-F. Tsai, H. Chen, Factors affecting text mining based stock prediction: text feature representations, machine learning models, and news platforms, Appl. Soft Comput. 130 (2022), 109673.

[32] K.K. Yun, S.W. Yoon, D. Won, Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process, Expert Syst. Appl. 186 (2021), 115716.

[33] M. Zolfaghari, S. Gholami, A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction, Expert Syst. Appl. 182 (2021), 115149.

[34] C. Wang, Y. Chen, S. Zhang, Q. Zhang, Stock market index prediction using deep Transformer model, Expert Syst. Appl. 208 (2022), 118128.

[35] G. Kwon, H. Lee, Time series KSTAR PF superconducting coil temperature forecasting using recurrent transformer model, Fusion Eng. Des. 193 (2023), 113693.

[36] X. Wang, H. Liu, J. Du, Z. Yang, X. Dong, CLformer: locally grouped auto-correlation and convolutional transformer for long-term multivariate time series forecasting, Eng. Appl. Artif. Intell. 121 (2023), 106042.

[37] B. Huang, K. Ruan, W. Yu, J. Xiao, R. Xie, J. Huang, ODformer: spatial–temporal transformers for long sequence Origin–Destination matrix forecasting against cross application scenario, Expert Syst. Appl. 222 (2023), 119835.

[38] M. Rashidpoor Toochaei, F. Moeini, Evaluating the performance of ensemble classifiers in stock returns prediction using effective features, Expert Syst. Appl. 213 (2023), 119186.

[39] H. Liu, C. Yu, H. Wu, Z. Duan, G. Yan, A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting, Energy 202 (2020), 117794.

[40] X. Liu, M. Qin, Y. He, X. Mi, C. Yu, A new multi-data-driven spatiotemporal PM2.5 forecasting model based on an ensemble graph reinforcement learning convolutional network, Atmos. Pollut. Res. 12 (10) (2021).

[41] Y. Chengqing, Y. Guangxi, Y. Chengming, Z. Yu, M. Xiwei, A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks, Energy 263 (2023), 126034.

[42] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C.-W. Qiu, J. Qiu, K. Hua, W. Su, J. Wu, H. Xu, Y. Han, C. Fu, Z. Yin, M. Liu, R. Roepman, S. Dietmann, M. Virta, F. Kengara, Z. Zhang, L. Zhang, T. Zhao, J. Dai, J. Yang, L. Lan, M. Luo, Z. Liu, T. An, B. Zhang, X. He, S. Cong, X. Liu, W. Zhang, J. P. Lewis, J.M. Tiedje, Q. Wang, Z. An, F. Wang, L. Zhang, T. Huang, C. Lu, Z. Cai, F. Wang, J. Zhang, Artificial intelligence: a powerful paradigm for scientific research, Innovation 2 (4) (2021), 100179.

[43] H. Wu, J. Xu, J. Wang, M.J.A.i.N.I.P.S. Long, Autoformer: Decomposition Transformers With Auto-Correlation For Long-Term Series Forecasting, 34, 2021, pp. 22419–22430.

[44] Y. Jiang, T. Gao, Y. Dai, R. Si, J. Hao, J. Zhang, D.W. Gao, Very short-term residential load forecasting based on deep-autoformer, Appl. Energy 328 (2022), 120120.

[45] N. Kitaev, Ł. Kaiser, A. Levskaya, Reformer: the efficient transformer, arXiv preprint (2020), 04451 arXiv:.

[46] J. Chen, H. Huang, H. Chen, Informer: irregular traffic detection for containerized microservices RPC in the real world, High-Conf. Com. 2 (2) (2022), 100050.

[47] M. Deb, B. Debbarma, A. Majumder, R. Banerjee, Performance –emission optimization of a diesel-hydrogen dual fuel operation: a NSGA II coupled TOPSIS MADM approach, Energy 117 (2016) 281–290.

[48] H. Liu, C. Yu, C. Yu, C. Chen, H. Wu, A novel axle temperature forecasting method based on decomposition, reinforcement learning optimization and neural network, Adv. Eng. Inf. 44 (2020), 101089.

[49] X. Mi, C. Yu, X. Liu, G. Yan, F. Yu, P. Shang, A dynamic ensemble deep deterministic policy gradient recursive network for spatiotemporal traffic speed forecasting in an urban road network, Digit. Signal Process. 129 (2022), 103643.

[50] M. Su, H. Liu, C. Yu, Z. Duan, A new crude oil futures forecasting method based on fusing quadratic forecasting with residual forecasting, Digit. Signal Process. 130 (2022), 103691.

[51] S. Verma, Forecasting volatility of crude oil futures using a GARCH–RNN hybrid approach, Intell. Syst. Account. Finance Manag. 28 (2) (2021) 130–142.