

Original Paper

Multivariable sales prediction for filling stations via GA improved BiLSTM

Shi-Yuan Pan, Qi Liao*, Yong-Tu Liang

Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum, Beijing, 102249, China

ARTICLE INFO

Article history:

Received 22 November 2021
Received in revised form
6 May 2022
Accepted 9 May 2022
Available online 13 May 2022

Edited by Xiu-Qiu Peng

Keywords:

Refined oil
Multivariable prediction
BiLSTM
Genetic algorithm
Future influence

ABSTRACT

Accurate sales prediction in filling stations is the basis to fill in the refined oil in time and avoid the out-of-stock as much as possible. Considering the defect of great “lag” in the general time series model, this paper summarizes the multiple factors that influence the oil sales and develops a multivariable long short-term memory (LSTM) neural network, with the hyper-parameters being improved by the genetic algorithm (GA). To further improve the prediction accuracy, the proposed LSTM neural network is generalized to bidirectional LSTM (BiLSTM), in which the impact of future factors on present sales can be taken into account by backward training. Finally, different LSTM structures and genetic algorithm parameters are tested to discuss their impact on prediction accuracy. Results demonstrated that genetic algorithm improved BiLSTM model is superior to extreme gradient boosting, ARIMA, and artificial neural network, having the highest accuracy of 89%.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

1.1. Background

With the continuous progress of global industrialization and urbanization, the demand for energy is growing more and more rapidly (Jiang and Lin, 2012). Energy is vital to the sustainable development of any country, whether it is social, economic or environmental. As an important part of primary energy, refined oil is regarded as a strategic commodity in the international community and enables global economic growth. It is widely used in automobiles, motorcycles, speedboats, helicopters, agricultural, forestry aircraft and other transportation tools, becoming an increasingly important role in daily life. As the critical urban infrastructure to provide energy supplies, filling station retail accounts for more than 70% of the total sales of refined oil.

The task of the supply chain is to transport refined oil products from refineries to oil depots by pipelines or trains and then to filling stations through trucks (Liang et al., 2012; Wang et al., 2021). Accurate prediction of energy demand can not only effectively capture the trend of energy demand, but it also helps suppliers make

accurate decisions (Suganthi and Samuel, 2012). Similarly, the accurate prediction of filling station sales is significant for the inventory management and distribution planning of petroleum enterprises since it provides the basis for replenishing the oil timely to avoid out-of-stock (Wang et al., 2015; Wei et al., 2021). However, most of the existing approaches only consider historical sales and time variables and deal with the time series data by using average value or linear models (Abdel-Aal and Al-Garni, 1997). Actually, the sales at filling stations are affected by multiple factors such as weather, week or holiday.

1.2. Related work

With the rapid development of science and technology, many forecasting methods have been put forward to improve the accuracy of prediction. Recent studies have comprehensively summarized energy prediction models proposed by predecessors (Suganthi and Samuel, 2012; Deb et al., 2016). The current mainstream prediction methods are as follows: time series model, regression-based model, grey prediction model, fuzzy logic method, BP neural network, support vector regression and LSTM neural network. Among time series models, ARIMA (Geurts et al., 1977) is the most common model. The characteristics and applied research fields of the above methods are summarized in Table 1:

* Corresponding author.

E-mail address: qliao@cup.edu.cn (Q. Liao).

Table 1
Summary of the above methods.

Methods	Data trend characteristics		Forecast period		The number of variables		Literature
	Linear	Nonlinear	Short term	Long term	Multivariate	Univariate	
Regression-based	✓		✓		✓		Sahraei et al. (2021) Bianco et al. (2009)
Time series	✓			✓		✓	Ciulla and D'Amico (2019) Yuan et al. (2016) Ediger and Akar (2007)
Grey prediction	✓			✓		✓	Wang et al. (2012) Huang et al. (2021a) Huang et al. (2021b)
Fuzzy logic		✓	✓		✓		Wang et al. (2018) Efendi et al. (2015)
BP neural network		✓	✓		✓		Sadaei et al. (2017) Torrini et al. (2016)
SVM	✓		✓			✓	Ekonomou (2010) Deb et al. (2016) Deb et al. (2015)
LSTM		✓	✓	✓	✓		Zhu et al. (2015) Xu et al. (2021) Ma et al. (2018) (Asala et al., 2017) Gupta and Pandey (2018) Asala et al. (2019) Chebeir et al. (2019) Wei et al. (2019) Laib et al. (2019) Lu et al. (2021)

Note: The symbol “✓” means the relative superiority of predictive performance.

The sales volume of filling stations which is affected by a variety of objective factors has obvious nonlinear characteristics. The predictive problem in this paper can be deal with a short-term prediction. Among the above methods, the BP neural network and the fuzzy logic model meet the prediction requirements of this problem, which is non-linear and affected by multiple factors. However, the fuzzy logic model is not applicable because the data of this problem are all determined and the BP neural network is easy to overfit.

Hochreiter and Schmidhuber (1997) was the first to propose LSTM neural network for processing and predicting important events with very long intervals and delays in time series. LSTM performs well in multi-feature regression due to its unique gating mechanism. It can automatically decide which information to remember or forget, so as to avoid the influence of redundant features or the low accuracy caused by over-fitting.

Nowadays, LSTM has been widely applied to address the limitations of traditional predictive methods in the field of energy forecasting. Tulensalo et al. (2020) used LSTM model to learn the long-term relationship of hourly time series data through three features which are electricity market, local weather and calendar and predict the total power system grid loss. Sagheer and Kotb (2019) proposed the deep LSTM(DLSTM) model, which avoids the limitations of shallow neural network architecture by superimposing more LSTM layers when predicting long-interval time series data sets. Additionally, genetic algorithm (GA) is applied to optimize the configuration of DLSTM architecture. Zheng et al. (2020) proposed a combined model based on LSTM and particle swarm optimization (PSO) for multi-regional solar power output prediction. In this paper, sensitivity analysis of the model is conducted and different LSTM structures are compared. Then, PSO is used for LSTM hyperparameter optimization. Moreover, results indicate that the proposed prediction model outperforms basic long short-term memory, artificial neural network and extreme gradient boosting. Laib et al. (2019) used the LSTM model to predict natural gas consumption in different regions of the country. It is

worthy to mention that the consumption of different regions of the country is firstly analyzed by clustering to reduce the non-stationarity of the time series. In addition, LSTM is used to predict the next day's gas consumption considering historical factors, meteorological factors (such as temperature, wind speed, humidity and sunshine) and economic factors (oil price, number of customers, GDP and gas price). A case study of natural gas consumption in Algeria proves the effectiveness of this method. More recently, Li and Becker (2021) proposed a hybrid model based on LSTM and feature selection algorithm for day-ahead electricity price prediction under consideration of market coupling. The importance of features was evaluated based on Shapley value before LSTM model which was used for electricity price prediction.

Since the sale data of filling stations has the features of great fluctuation and nonlinearity, the performance of traditional time-series models such as ARIMA become poor and cannot meet the accuracy requirements of forecasting. At present, the global industry is experiencing the wave of artificial intelligence, and the petroleum industry is also undergoing intelligent transformation. LSTM model, as one of the best time series prediction models in the field of artificial intelligence of deep learning, can effectively deal with nonlinear time series prediction by adding other features to assist prediction. At the same time, different LSTM structures are compared to select the optimal LSTM structure. Moreover, we consider the influence of future factors on present in the prediction model by using BiLSTM. Finally, the optimal hyperparameters cannot be obtained by manual hyperparameter tuning due to the large number of hyperparameters in LSTM model, so GA is used to automate hyperparameter tuning to further improve the accuracy of the model.

1.3. Contributions and paper organization

The contributions of this paper to the field of filling station sales forecasting are the following:

- (1) By introducing the LSTM model, this paper solves the low prediction accuracy problem of traditional time series model such as ARIMA. In addition, the LSTM model can effectively reduce the effect of lag caused by excessive fluctuation by adding the objective factors that affect the sales of filling stations to assist prediction.
- (2) The BiLSTM model can further improve accuracy because the influence of future on present can be taken into account in the backward training.
- (3) Due to the large number of hyperparameters in LSTM model, the optimal hyperparameters often cannot be obtained by manual hyperparameters tuning. For this reason, GA is used to automate hyperparameters tuning to further improve the accuracy of the model.
- (4) In this paper, the sales volume of three years of a filling station in Kunming, China is taken as a data case. We compare the proposed model with other algorithms to prove the superiority of it. The results show that its performance is the best and accuracy can up to 89%.

The rest of the paper is structured as follows: Section 2 describes the characteristics of the problems solved in this paper. Section 3 introduces the processing methods of text variables and the principles of GA and LSTM. Section 4 is the experimental part, which respectively introduces feature extraction, advantages of Embedding on dimension reduction of word vector, the influence of different LSTM structures on prediction accuracy and comparison between the proposed model and other algorithms. Section 5 gives the conclusions and suggestions for future research.

2. Problem description

The prediction of filling station sales is an indispensable part of the secondary logistics transportation of refined oil products. Accurate prediction can help suppliers make efficient distribution plans for the next day. However, the prediction of sales volume has the following difficulties:

- (1) The sales volume of filling stations fluctuates greatly and has obvious nonlinear characteristics, which increases the difficulty of prediction. Therefore, the characteristic of large fluctuation will make a prediction have the defect of great “lag”, which leads to part of the prediction results need to be manually adjusted according to experience.
- (2) The sales forecasting problem of filling stations is a short-term forecasting problem. It only needs to forecast the sales of the next day.
- (3) The sales of filling stations are affected by a variety of objective factors, such as holidays, weeks, weather, temperature, oil prices, etc. For example, due to the Spring Festival, the sales before the Spring Festival period are much higher than normal sales, while sales during the Spring Festival are much lower than normal sales. For another example, weekdays from Monday to Friday and weekends will affect people's travel plans and thus affect the sales of filling stations. In addition, the extreme weather affects people's daily travel and lead to a decline in sales.
- (4) The periodicity of filling station sales is not obvious.
- (5) Future factors have an impact on the present. For example, if people have holiday travel plans tomorrow, they will choose to fuel up the day before the trip.

In case of these difficulties, the prediction model should be multivariate, and the goal of adding objective factors which affect sales volume is to capture the relationship between sales volume

and these objective factors. However, various methods among the multivariate prediction methods, such as BP neural network or SVM, are easy to over-fit because of the high feature dimensions. On the contrary, it is widely demonstrated that LSTM is suitable for processing and predicting important events with very long intervals and delays in time series and performs well in the prediction of higher dimensions. Therefore, LSTM model is used in this paper, and the characteristics of holidays, weeks, weather, temperature and oil price are added to assist in predicting the sales of filling stations. Finally, GA is used to find the optimal hyperparameters of LSTM to further improve the model accuracy.

The heat map of Fig. 1 will be explained below. Weather: red for sunny days, orange for cloudy days, yellow for cloudy days, green for rainy days, and the darker the green, the heavier the rain. Oil price: the darker the red, the higher the oil price, and the darker the green, the lower the oil price. Holidays: green represents normal days and red represents holidays. Temperature: the darker the red, the higher the temperature, and the darker the green, the lower the temperature.

Meanwhile, in Section 4.3, this paper compares the influence of different LSTM structures on the prediction accuracy and finds that bidirectional LSTM is superior to unidirectional LSTM in the prediction work. The explanation of this result is as follows. As shown in Fig. 1, it is assumed that the historical sales of the first seven days and other auxiliary features are used to predict the sales of the next day. The characteristic vectors of the first seven days are set respectively $x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$. We assume that the weather and week conditions are shown as Fig. 1. Since the bidirectional LSTM model has the characteristics of forward and backward training, in backward training, it can learn the rules that cannot be learned by forward training. When people learn from the weather forecast that the weather is bad tomorrow, they will choose to refuel on that day. Or when they have holiday travel plans tomorrow, they will choose to refuel on the day before the trip. Bidirectional LSTM can learn these rules in backward training, but unidirectional LSTM cannot. For example in Fig. 1, in backward training from day x_{t-4} to day x_{t-5} , it is a thunderstorm on day x_{t-4} . If the sales volume of day x_{t-5} is higher than that of day x_{t-4} , BiLSTM could learn the above rules. Similarly, in backward training from day x_{t-2} to day x_{t-3} , if day x_{t-2} is a holiday and sale volume of day x_{t-3} is higher than that of day x_{t-2} , BiLSTM can also learn the above rules.

3. Methodology

In this paper, the data is daily and a total of 940 days' data of a filling station in Kunming from 2019 to 2021 are selected, with 70% as training set and 30% as test set. Firstly, this paper carries on the feature extraction. We select the objective factors, which affect filling station sales more to make auxiliary projections. Then, we vectorize the text-type factor. In other words, convert text-type factor variables to word vectors. In addition, this paper compares different LSTM structures and use GA to optimize the hyperparameters. Finally, the GA-LSTM model with the best performance is also compared with other algorithms to prove the effectiveness of the proposed method. The flowchart of this paper is shown in Fig. 2:

3.1. Word to vector

The weather and week auxiliary features of this paper are text-type variables that need to be converted into word vectors. One-hot encoding is carried out for text encoding of weather and week, and dimension reduction is carried out using Embedding layer. Moreover, the dimension is reduced through the weight matrix

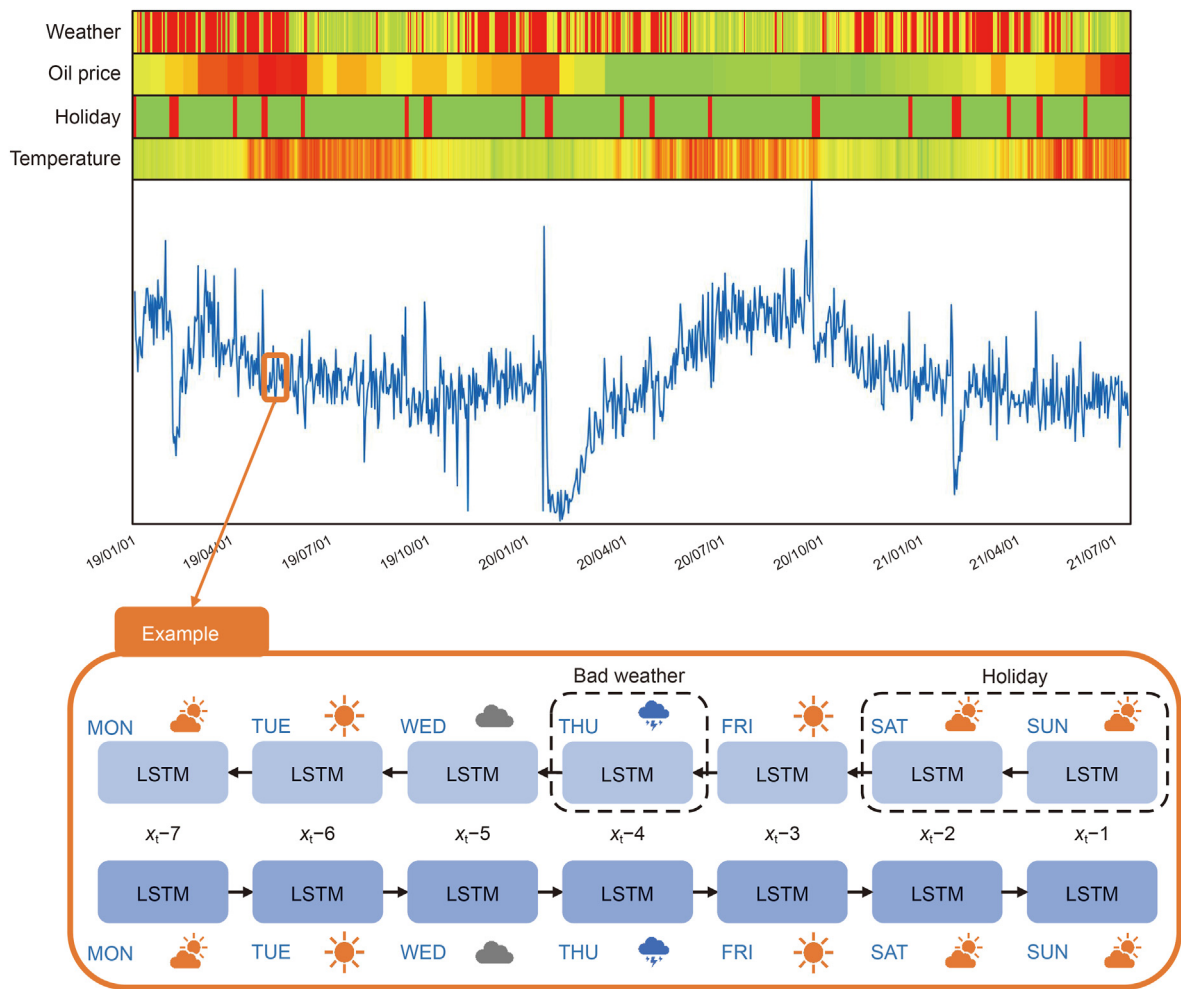


Fig. 1. Multivariate prediction with objective factors.

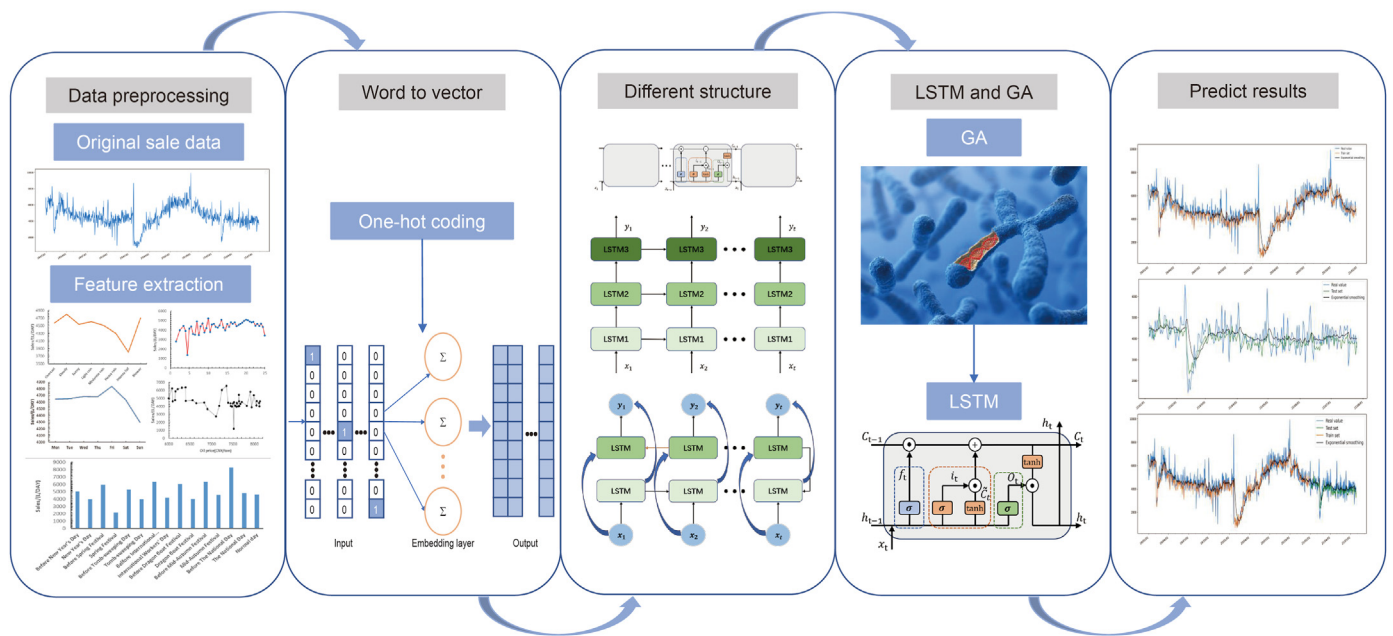


Fig. 2. Flowchart.

calculation of Embedding layer. Then, the Embedding matrix assigns a fixed length vector to each word and the length of vectors can be set by ourself. The principle of the embedding layer is shown in Fig. 3:

Suppose there are m weather texts with n weather types, and x week texts with y week types. Then, the one-hot encoding vectors of each weather text q_a and each week text p_b are respectively $q_a = [q_{a1}, q_{a2}, \dots, q_{an}]$, $a = 1, 2, \dots, m$; $p_b = [p_{b1}, p_{b2}, \dots, p_{by}]$, $b = 1, 2, \dots, x$. Finally, the weather vector matrix Q and the week vector matrix P are as follows:

$$Q = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \dots & q_{mn} \end{bmatrix}, P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1y} \\ p_{21} & p_{22} & \dots & p_{2y} \\ \vdots & \vdots & \ddots & \vdots \\ p_{x1} & p_{x2} & \dots & p_{xy} \end{bmatrix}, Q \in R^{m \times n}, P \in R^{x \times y} \quad (1)$$

Embedding layer which needs to learn weight matrix is essentially a neural network with a hidden layer. If the dimension of weather vector matrix Q is reduced from $m \times n$ to $m \times k$, that is, m vectors whose dimension is $1 \times n$ are reduced to m vectors whose dimension is $1 \times k$, and the weight matrix W_1 is:

$$W_1 = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{bmatrix}, W_1 \in R^{n \times k} \quad (2)$$

Similarly, to reduce the dimension of week vector matrix P from dimension $x \times y$ to $x \times h$, the weight matrix W_2 is:

$$W_2 = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1h} \\ b_{21} & b_{22} & \dots & b_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ b_{x1} & b_{x2} & \dots & b_{xh} \end{bmatrix}, W_2 \in R^{y \times h} \quad (3)$$

The new weather vector matrix Q' and the new week vector matrix P' obtained by Embedding layer for dimension reduction are:

$$Q' = Q \times W_1 = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1k} \\ q_{21} & q_{22} & \dots & q_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \dots & q_{mk} \end{bmatrix}, Q' \in R^{m \times k} \quad (4)$$

$$P' = P \times W_2 = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1h} \\ p_{21} & p_{22} & \dots & p_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ p_{x1} & p_{x2} & \dots & p_{xh} \end{bmatrix}, P' \in R^{x \times h} \quad (5)$$

In this paper, the week feature is reduced from 7 dimensions to 1 dimension through the Embedding layer, while the weather feature is reduced from 8 dimensions to 2 dimensions.

3.2. LSTM model

The traditional Recurrent Neural Network (RNN) is one of the recursive neural network models that can be applied for modeling of sequential data, but it will be difficult for RNN to transmit information from the earlier time step to the later time step. To address this issue, LSTM was proposed to solve the short-term memory problem of RNN. It has an internal mechanism called “gate”. The “gate” structure will learn which information should be saved or forgotten during training. It can then make predictions by passing relevant information along long sequences.

(1) LSTM cell

A cell unit in the LSTM model contains three gate structures: forgetting gate, input gate and output gate, as well as two states: cell state C_t and hidden layer state h_t . The hidden layer state is the output of the final network, while the cell state will participate in the calculation of the hidden layer state. The structure of each cell unit is shown in Fig. 4:

Assuming that the data of previous r days are used to predict the sales volume of the next day, then, the input vector x_t of LSTM at time t is:

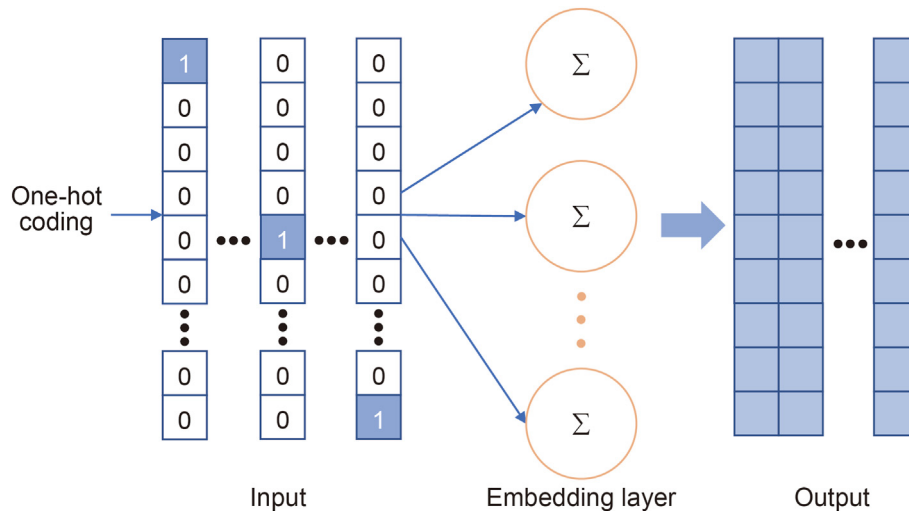


Fig. 3. Word to vector.

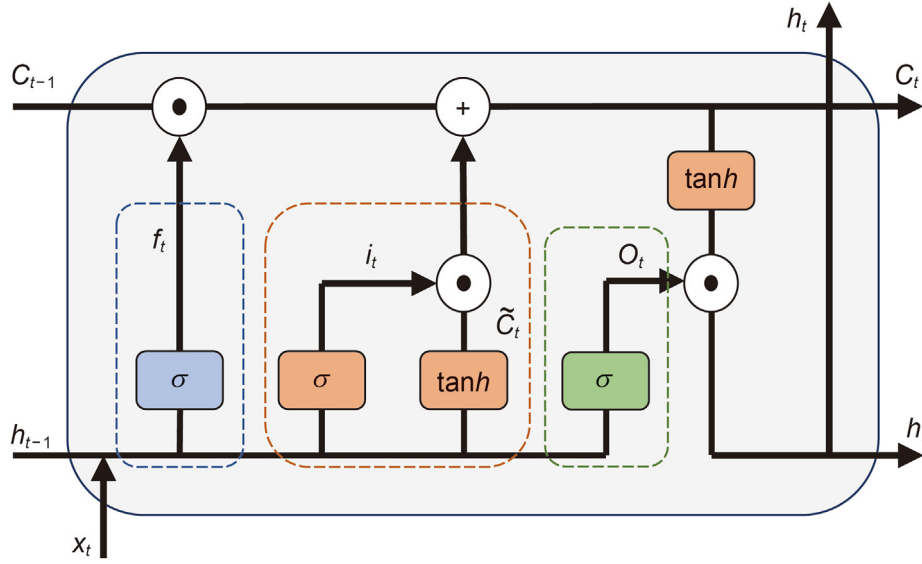


Fig. 4. Structure of LSTM (Hochreiter and Schmidhuber, 1997).

$$x_t = [d_{t-r}, d_{t-(r-1)}, \dots, d_{t-1}]$$

be -1 and 1 by $\tanh$ function. Then the cell state vector c_{t-1} at time step $t-1$ is calculated with these three vectors to obtain the cell state vector C_t at time step t . Finally, through Eq. (11), the output gate vector o_t and the cell state vector C_t are combined to obtain the state output h_t of the hidden layer at time step t .

$$d_{t-k} = [\text{sale}_{t-k}, \text{weather}_{t-k}, \text{week}_{t-k}, \text{temperature}_{t-k}, \text{oil_price}_{t-k}], k = 1, 2, \dots, r$$

d_{t-k} is the data of the last k days at time t . sale_{t-k} , weather_{t-k} , week_{t-k} , temperature_{t-k} , oil_price_{t-k} represent the sales volume, weather, week, temperature, oil price of the last k days at time step t . When the input vector of LSTM at time step t is x_t , the output h_t of LSTM's hidden layer state is $h_t = [h_1, h_2, \dots, h_r]$, and the cell calculation process of each time step is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) + b_f \quad (6)$$

$$l_t = \sigma(W_l \cdot [h_{t-1}, x_t]) + b_l \quad (7)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (8)$$

$$C_t = f_t \cdot C_{t-1} + l_t \cdot \tilde{C}_t \quad (9)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t]) + b_o \quad (10)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (11)$$

W_f , W_l , W_C , W_o are weight matrices that are automatically learned and updated during training. b_f , b_l , b_C , b_o are offsets. At time step t , the input of the cell unit is the cell state C_{t-1} at time step $t-1$, the hidden layer state h_{t-1} at time step $t-1$ and the new input x_t . The output vector f_t of Eq. (6) in the forgetting gate and the output vector l_t of Eq. (7) in the input gate are pushed the value to be between 0 and 1 by Sigmoid function respectively, and the output vector $\tilde{C}_t$ of Eq. (8) in the input gate are pushed the value to

(2) Single-layer LSTM

The single-layer LSTM takes a chain form of repeating neural network (RNN) modules. The input of single-layer LSTM is step by step according to time series, and the cell state is updated in each time step. At time step t , the cell state of unit C_{t-1} represents the input state of cell unit remembering at time step $t-1$ and previous time step, and h_{t-1} represents the output of cell unit at time step $t-1$. Therefore, when multiple cell units are connected together, a complete single-layer LSTM is formed as shown in Fig. 5. In single-layer LSTM, the input sequence is $x_1, x_2, \dots, x_t$, and each of them is a vector containing n features ($x_t = x_{t1}, x_{t2}, \dots, x_{tn}$). At this time step, LSTM output is h_t , and then a new sequence is input for the next round of prediction.

(3) Multi-layer LSTM

Multi-layer LSTM is slightly different from single-layer LSTM. The structure of multi-layer LSTM resembles artificial neural network and has multiple hidden layers, while the single-layer LSTM has only one hidden layer (see Fig. 6). In the case of large data volume, it is widely demonstrated that multi-layer LSTM has higher accuracy than single-layer LSTM. In multi-layer LSTM, the output of the current time step is the input of the next layer at the same time step, and the output of the current time step is the output of the next time step at the same hidden layer.

(4) Bi-directional LSTM

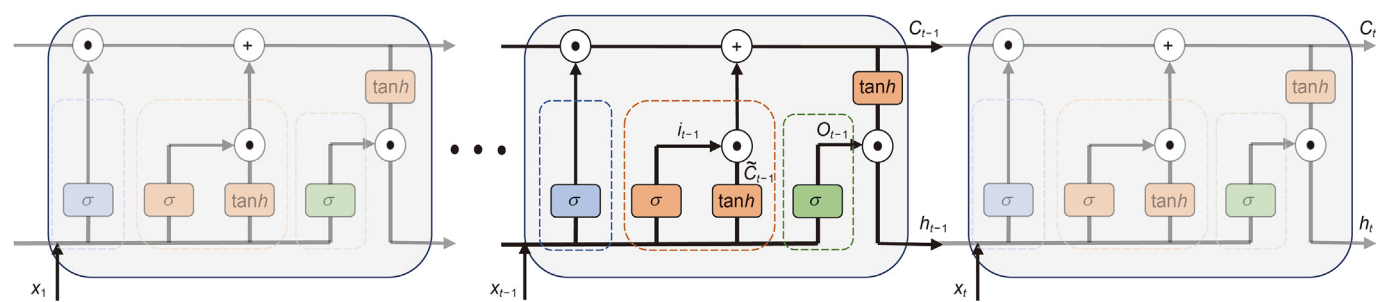


Fig. 5. Single-layer LSTM (Hochreiter and Schmidhuber, 1997).

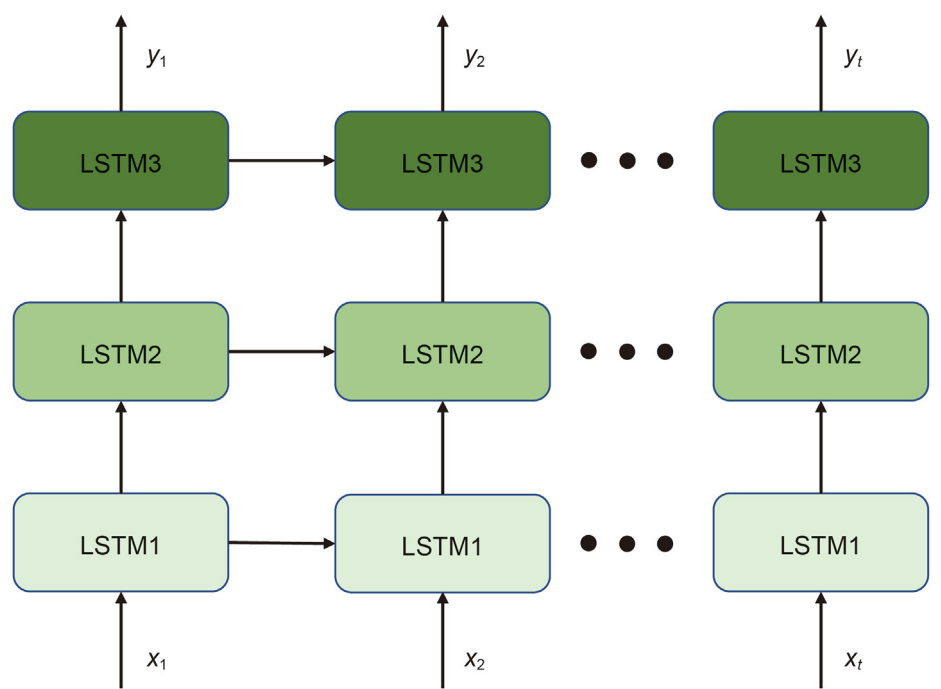


Fig. 6. Multi-layer LSTM

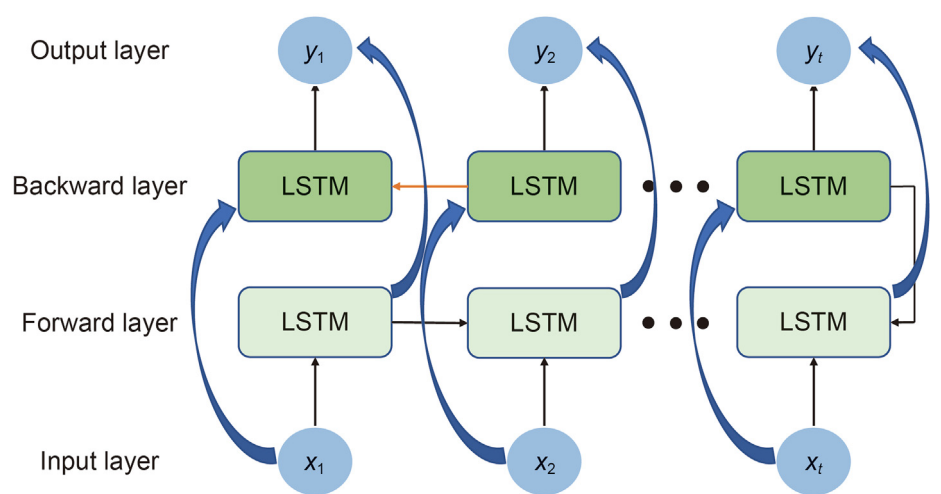


Fig. 7. BiLSTM.

Bidirectional LSTM (BiLSTM) network is an extended structure of ordinary LSTM (Schuster and Paliwal, 1997). BiLSTM's forward and backward training which are connected to an output layer are sequential LSTM models (see Fig. 7). In the forward layer, LSTM is applied to the input sequence. In the backward layer, the backward form of the input sequence is fed into the LSTM model. Applying LSTM twice can improve the long-term dependence of learning and thus improve the accuracy of the model. Meanwhile, Siarni-Namini et al. (2019) compared the performance of ordinary LSTM and BiLSTM in time series prediction, and the results showed that BiLSTM-based model provided a better prediction compared with the conventional LSTM-based model.

3.3. Genetic algorithm

Holland (1975) proposed GA firstly. Inspired by biological evolution, GA has good performance in finding global optimal solutions and can find global optimal solutions for non-differentiable problems. This paper uses GA to find the optimal hyperparameters of the LSTM model. Each characteristic parameter uses the corresponding chromosome encoding mode, and the chromosome composition of all populations is the same. After that, the next generation population is generated by crossover, mutation and selection operations. By the crossover operation, genes from the same parts of chromosomes pairs are exchanged to explore a larger solution space and obtain a better solution more easily. Then, the genetic information is changed by random replacement of genes in chromosomes through mutation operation to prevent falling into local optimal solution. In the selection operation, the fitness function is used to evaluate the adaptability of an individual. The stronger the adaptability, the greater the probability of the individual being selected.

In this paper, GA encoding is binary encoding. Each individual contains five characteristic parameters: number of neural units, L2 regularization weights, dropout probability, epoch, batch size. When the population is initialized, the population has n individuals, and the chromosome composition of each individual N_i is:

$$N_i = [a_i, b_i, c_i, d_i, e_i], i = 1, 2, \dots, n \quad (12)$$

a_i, b_i, c_i, d_i, e_i are respectively: number of neural units, L2 regularization weights, dropout probability, epoch, batch size. Their value ranges are as follows:

$$a_i \in [50, 150], b_i \in [0.0001, 0.01], c_i \in [0.2, 0.8], d_i \in [300, 800], e_i \in [16, 128] \quad (13)$$

The fitness function of the GA considers root mean squared error (RMSE) and mean absolute error (MAE) respectively for multi-objective minimization optimization, that is, the fitness function f of the GA is:

$$\min f = RMSE + MAE \quad (14)$$

We assume the fitness function value of each individual is $f_i, i = 1, 2, \dots, n$, and the probability of the individual being selected is defined as $p_i = \frac{f_i}{\sum_{k=1}^n f_k}$. Then, the optimal solution can be found by continuously performing selection, crossover and mutation operations in Table 2.

4. Experiments and analysis

4.1. Evaluation indexes and feature extraction

The evaluation indexes selected in this paper are root mean squared error (RMSE), mean absolute error (MAE) and accuracy (ACC). RMSE is the square root of the ratio of the square deviation between the predicted value and the real value. MAE is the average value of the absolute error between the observed value and the real value. ACC is the percentage of the absolute value between the predicted value and the real value. The three standard formulas are as follows (y_p indicates the predicted value, and $\hat{y}_p$ indicates the actual value):

$$RMSE = \sqrt{\frac{1}{q} \sum_{p=1}^q (y_p - \hat{y}_p)^2} \quad (15)$$

$$MAE = \frac{1}{q} \sum_{p=1}^q |y_p - \hat{y}_p| \quad (16)$$

$$ACC = \frac{1}{q} \sum_{p=1}^q \frac{|y_p - \hat{y}_p|}{\hat{y}_p} \quad (17)$$

The following features are added from the raw data to assist forecast: week, temperature, weather, holidays, and oil prices. For the above features, the sales volume during the holidays and the day before the holidays (the average sales volume of the previous two days during the Spring Festival and National Day) is selected to calculate the average sales, and the relationship diagram with sales volume is shown in Fig. 8a–e:

According to Fig. 8a–e, the conclusions are as follows:

- (1) Take the average of the same oil price to observe the impact of oil price on sales volume. Although the oil price is affected by COVID-19, it can still be seen from the chart that the sales volume tends to decline slightly as the oil price rises. There is a correlation between gas station sales and oil prices.
- (2) Take the average value of the same temperature to observe the impact of temperature on sales volume. Due to regional relations, the temperature in Kunming fluctuates within the range of 1.5 °C–25 °C, and low or high temperature rarely occurs. Sales volume has a certain correlation with temperature.

Table 2

The GA for parameters optimization.

Input: the number of generations G , the population of each generation n , crossover probability p_1 , mutation probability p_2
Population initialization: initializes the chromosomes of each individual in the first generation $\{N_{i,0}\}_{i=1,2,\dots,n}$, $N_i = [a_i, b_i, c_i, d_i, e_i]$, $i = 1, 2, \dots, n$
for $g = 0, 1, \dots, G - 1$:
Evaluation: calculate the fitness function of each individual: $\{f_{i,g}\}_{i=1,2,\dots,n}$
Crossover: individual $\{N_{i,g}\}_{i=1,2,\dots,n}$ and individual $\{N_{j,g}\}_{j=1,2,\dots,n}$ exchange their chromosomes to get $\{N_{ij,g+1}\}_{ij=1,2,\dots,n}$
Mutation: some genes of the offspring $\{N_{ij,g+1}\}$ obtained from the parent generation mutate
Selection: select the individual from $\{N_{ij,g+1}\}_{ij=1,2,\dots,n}$ by their fitness function value $\{f_{ij,g}\}_{ij=1,2,\dots,n}$ to generate $\{N_{i,g+1}\}_{i=1,2,\dots,n}$
End
Output: the final generation $\{N_{i,G}\}_{i=1,2,\dots,n}$ and their fitness value $\{f_{i,G}\}_{i=1,2,\dots,n}$

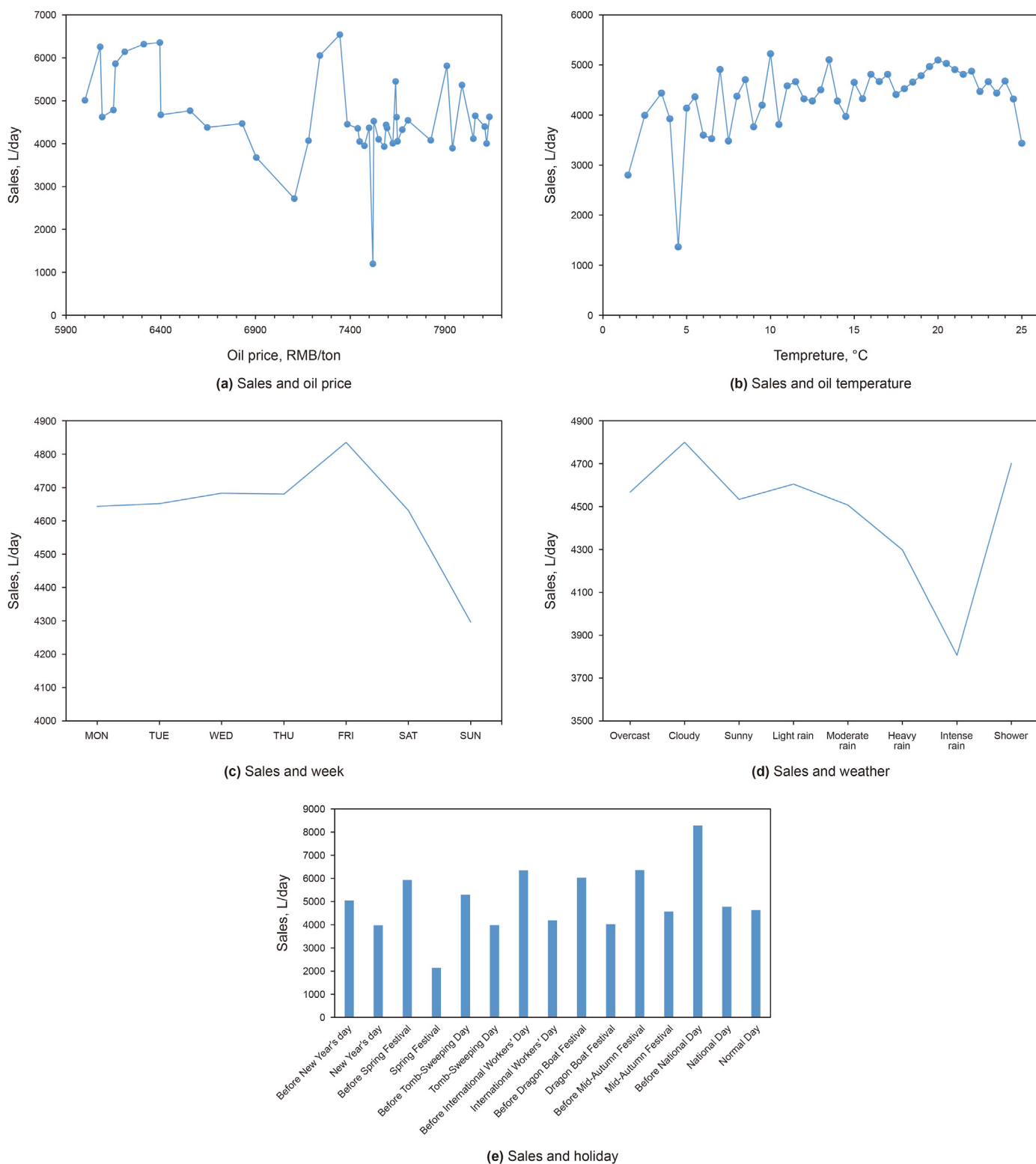


Fig. 8. The relationship between sales and different factors.

(3) By adding and averaging all sales volumes from Monday to Sunday, sales volume shows an upward trend that it reaches its peak on Friday, then, declines on Saturday and Sunday and reaches its lowest point on Sunday. Filling station sales have a strong correlation with the week.

(4) Take the average value of the same weather condition to observe the impact of weather on sales. Due to regional relations, the local weather does not include all weather types, but it can still be seen that heavy rain and other bad weather have a great impact on sales.

Table 3
Comparison of different combinations of features (test set).

Different combination of features		RMSE	MAE	ACC
Case 1	Historical sales, week, weather, holiday, temperature, oil price	588.4	445.9	88.3%
Case 2	Historical sales, week, weather, holiday, oil price	601.3	458.4	87.8%
Case 3	Historical sales, week, weather, holiday, temperature	595.7	449.4	88.1%
Case 4	Historical sales, week, weather, holiday	597.7	449.8	88.1%
Case 5	Historical sales	672.0	484.2	86.9%

- (5) Take the average value of holidays and the day before holidays to observe the impact of holidays on sales volume. Through the comparison, it is found that the sales volume before the holidays are higher than the usual sales volume. During the holidays, except for the Spring Festival, the sales volume is slightly higher than the usual sales volume. During the Spring Festival, the sales volume is much lower than the usual sales volume. There is a strong correlation between the filling station sales and holidays.

The correlation between sales volume and each auxiliary feature has been briefly analyzed above according to Fig. 8a–e. In order to further study the influence of features whether have a certain correlation with sales volume on prediction accuracy or not, features with different combinations are selected, and the results obtained are shown in Table 3. The results of Case 1, Case 2, Case 3, Case 4 are all superior to Case 5, which is univariate prediction. It is indicated that the addition of objective factor features for prediction can significantly improve the prediction accuracy of the model. At the same time, by comparing Case 1, Case 2, Case 3, Case 4, it can be seen that RMSE, MAE and ACC of Case1 are all optimal. It is indicated that adding temperature and oil price for prediction can still improve the model accuracy slightly.

4.2. Feature dimension reduction

Since the weather and week variables are text-type, it is necessary to convert text variables into word vectors. The text encoding adopts one-hot encoding. However, one-hot encoding leads to excessive feature dimensions and overfitting of the model. Therefore, this paper adopts Embedding layer to reduce feature dimensions of high-dimensional text to increase the accuracy of the model. The following two cases are compared: one-hot encoding only and one-hot encoding followed by Embedding for dimension reduction. Unidirectional LSTM and bidirectional LSTM models are respectively used for comparison, and the results obtained are shown in Table 4 and Table 5. It can be seen that the accuracy of the model with one-hot encoding and dimension reduction is obviously better than that with one-hot encoding only. At the same time, in the case of the one-hot coding model, the results show that the evaluation indexes' difference between the training set and the test set is too large, and the model overfits due to high feature dimensions. Therefore, Embedding method can effectively avoid model overfitting and improve the model's accuracy and generalization ability.

Table 4
Comparison of Feature dimension reduction in unidirectional LSTM.

	RMSE		MAE		ACC	
	Train set	Test set	Train set	Test set	Train set	Test set
One-hot and embedding	576.8	587.6	399.1	446.6	90.0%	88.3%
One-hot	333.6	628.1	210.7	483.4	94.3%	85.5%

4.3. LSTM structure

Different LSTM structures have different precision for different problems. In order to find the optimal model structure, this paper compares four different LSTM structures: single-layer LSTM, multi-layer LSTM, single-layer bidirectional LSTM (BiLSTM) and multi-layer bidirectional LSTM (BiLSTM). According to the empirical method and the scale of the problem in this paper, it can be roughly determined that the Cell units are between 60 and 100 and the number of LSTM layers does not exceed two layers. To avoid statistical uncertainty, we run each example 20 times for each case. LSTM parameters of dropout probability, loss, epochs, batch size, time step are 0.5, 'MSE', 500, 64, 7 respectively. The comparison results are shown in Table 6. It can be seen from Table 6 that the accuracy of single-layer LSTM model is better than that of multi-layer LSTM, and the precision of single-layer bidirectional LSTM model is better than that of single-layer unidirectional LSTM. In addition, according to the standard deviation of Table 6, it shows that the stability of single-layer bidirectional LSTM is slightly better than that of single-layer unidirectional LSTM in most cases. The reason is that bidirectional LSTM will be trained separately from forward and backward. If the weather is bad tomorrow or there are holiday travel plans tomorrow, the backward training in bidirectional LSTM can effectively capture this information to update the prediction strategy. The single-layer bidirectional LSTM model is finally selected for the following reasons: (1) due to the small scale of the problem, the single-layer LSTM model has higher accuracy; (2) The single-layer LSTM model has a shorter training time; (3) In the time series prediction problem, the bidirectional LSTM model usually has higher accuracy, mainly because the bidirectional LSTM model can consider the influence of future factors that unidirectional LSTM cannot during the backward training process.

4.4. GA-BiLSTM

The hyperparameters are parameters which are set before the start of the learning process rather than obtained through training. In general, different hyperparameter settings will lead to different results, so it is necessary to optimize the hyperparameters and select a group of optimal hyperparameters for the model to improve the performance. The hyperparameters involved in the BiLSTM model include the number of neural units, L2 regularization weights, dropout probability, epoch and batch size. With the sum of RMSE and MAE of the model used as fitness function for multi-objective optimization, the optimal hyperparameters are sought by GA. Simultaneously, crossover probability, generation and

Table 5
Comparison of Feature dimension reduction in bidirectional LSTM (BiLSTM).

	RMSE		MAE		ACC	
	Train set	Test set	Train set	Test set	Train set	Test set
One-hot and embedding	580.1	576.6	402.0	441.1	90.1%	88.4%
One-hot	321.7	672.8	197.3	519.2	95.2%	85.7%

Table 6
Different structures of LSTM.

	Cell units (1st layer)	Cell units (2nd layer)	RMSE's mean	RMSE's standard deviation	MAE's mean	MAE's standard deviation	ACC's mean	ACC's standard deviation
Single-layer LSTM 60	\		588.60	8.28	446.70	7.76	88.54%	0.11
Single-layer LSTM 80	\		590.92	11.52	449.74	9.52	88.40%	0.11
Single-layer LSTM 100	\		591.28	10.85	450.85	11.36	88.51%	0.11
Single-layer BiLSTM 60	\		587.50	10.33	444.15	8.91	88.53%	0.11
Single-layer BiLSTM 80	\		585.58	9.23	446.78	8.86	88.52%	0.08
Single-layer BiLSTM 100	\		585.37	10.83	450.07	10.67	88.54%	0.09
Multi-layer LSTM 60	30		598.45	13.03	457.85	11.78	88.40%	0.17
Multi-layer LSTM 80	40		600.00	15.93	454.91	14.65	88.35%	0.29
Multi-layer LSTM 100	50		600.77	12.03	456.12	10.09	88.38%	0.14
Multi-layer BiLSTM 60	30		597.54	12.01	454.10	11.06	88.41%	0.12
Multi-layer BiLSTM 80	40		594.22	12.58	453.98	11.12	88.35%	0.14
Multi-layer BiLSTM 100	50		604.32	12.58	458.73	11.12	88.40%	0.15

Table 7
Different GA parameter combination.

	Crossover probability	Generation	Population size	MAE + RMSE	MAE	RMSE
Case 1	0.3	15	15	1005.6	435.1	570.5
Case 2	0.3	30	20	1001.7	433.1	568.6
Case 3	0.3	20	30	1010.6	437.1	573.5
Case 4	0.3	20	20	1006.8	436.5	570.3
Case 5	0.7	15	15	1011.2	437.2	574.0
Case 6	0.7	30	20	1007.7	436.5	571.2
Case 7	0.7	20	30	1006.0	436.7	569.3
Case 8	0.7	15	15	1007.1	436.6	570.5
Basic LSTM	\	\	\	1034.2	446.6	587.6

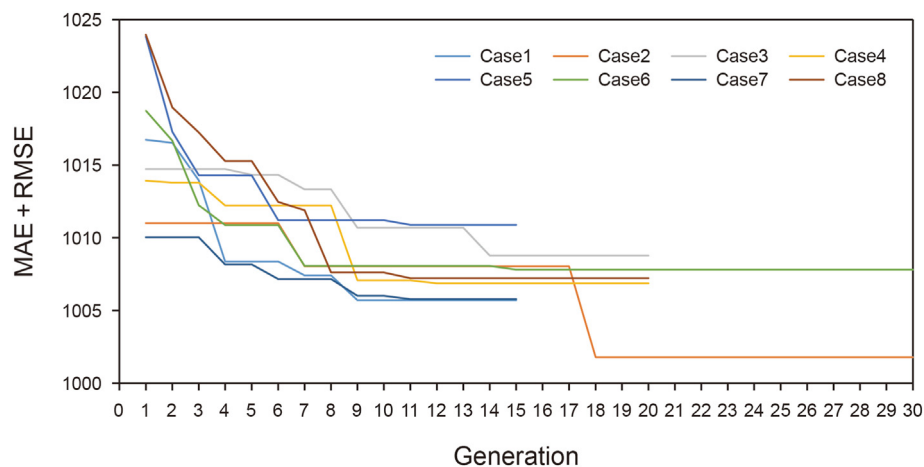


Fig. 9. MAE + RMSE of GA-BiLSTM.

population size of the GA have a certain influence on the results. For example, selecting the appropriate crossover probability can

determine whether the global optimal solution can be obtained. Table 7 shows the results obtained by using GA with different

Table 8
Comparison of different models.

	RMSE		MAE		ACC	
	Train set	Test set	Train set	Test set	Train set	Test set
BP neural network	639.7	628.8	479.9	491.9	87.7%	87.0%
XGBOOST	635.5	645.1	469.2	460.0	88.2%	87.2%
Exponential smoothing	\	764.3	\	551.2	\	85.1%
ARIMA	\	737.9	\	577.2	\	86.1%
Basic BiLSTM	576.8	587.7	399.1	446.6	90.0%	88.3%
GA-BiLSTM	571.2	569.3	422.1	436.7	90.6%	88.7%

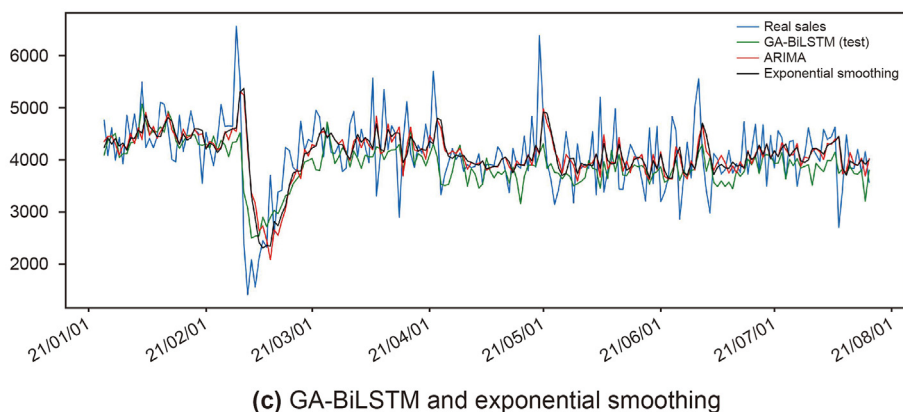
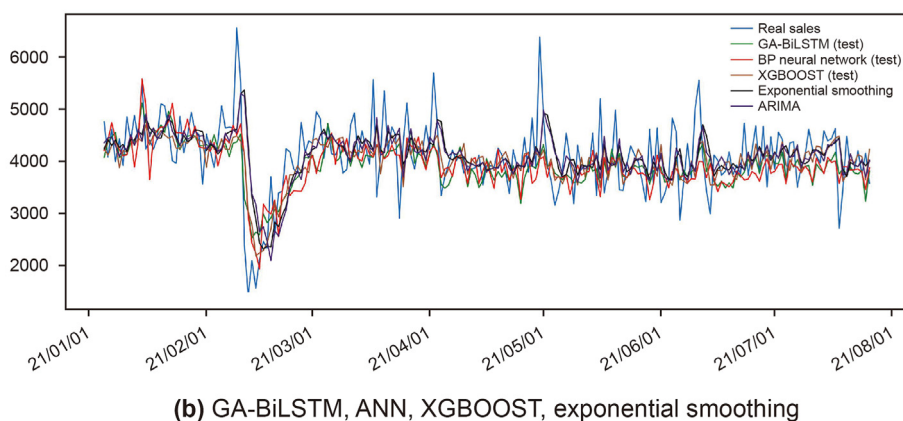
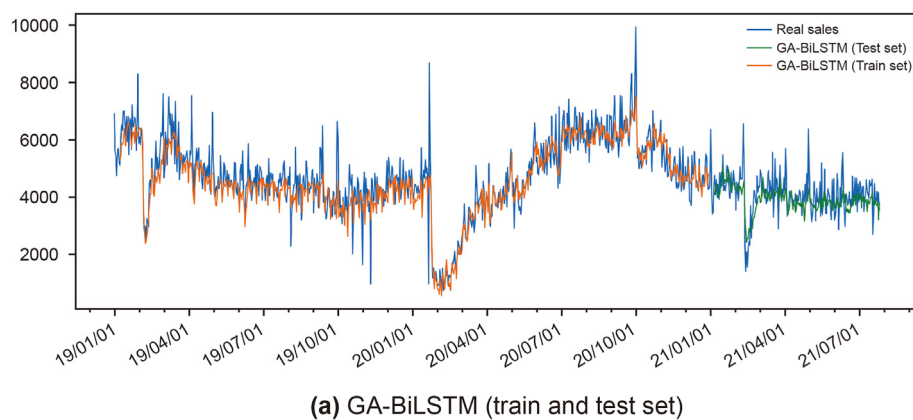


Fig. 10. The predicted results of different prediction models.

crossover probability, population number and individual number. Fig. 9 shows the convergence of fitness function with different GA

parameter combinations. As can be seen from Table 7, compared with Basic LSTM based on manual experience, GA-BiLSTM

effectively reduces the model error by 3%, and the result obtained when the crossover probability is 0.3 is better than that when the crossover probability is 0.7. Fig. 9 shows the convergence of eight different GA parameter combinations, and it can be seen from the figure that they all converge. Finally, according to the comparison, the number of neural units, L2 regularization weights, dropout probability, epoch, and batch size of optimal BiLSTM hyperparameters obtained are 83, 0.000685, 0.543, 456, 54.

4.5. Results and comparison

The structure of LSTM model and optimal parameters of BiLSTM have been determined in Section 4.3 and Section 4.4. To illustrate our superiority, the proposed GA-BiLSTM method is compared with basic BiLSTM, XGBOOST (Chen and Guestrin, 2016), ARIMA, BP neural network and exponential smoothing model which is currently used in business. Each model uses the last 30% of the data as a test set to test the accuracy of the prediction model. Since other algorithms have fewer hyperparameters, we use grid search to find their optimal parameters, and the results are as follows: (1) The weight coefficient α of the first-order exponential smoothing method is 0.35. (2) The number of neurons in the first layer of the BP neural network is 20, and the number of neurons in the second layer is 1. (3) The maximum number of iterations of XGBOOST is 22, the maximum depth is 3, and the learning rate is 0.1. The obtained prediction results are shown in Table 8. It shows that the results of the BiLSTM model are better than other models. In order to make the prediction results more intuitive, Fig. 10a–c shows the prediction results of different models. Fig. 10a is the comparison among the training set, test set of GA-BiLSTM model and real value. Fig. 10b is the comparison between the results of all test sets of the above methods and the real value. Fig. 10c is the comparison among the GA-BiLSTM model, ARIMA and the exponential smoothing model used in business. As shown in Fig. 10c, in places with large fluctuations, for example, the sales volume fluctuates greatly around the Spring Festival days. Compared with ARIMA and the exponential smoothing model, GA-BiLSTM has the best performance and the part outlined by the dotted line in Fig. 10c can show that GA-BiLSTM model can reduce the effect of lag caused by excessive fluctuation. In other words, when the sales are with large fluctuations, GA-BiLSTM model's prediction can react more quickly. Meanwhile, Fig. 10c compares the results among the exponential smoothing model, ARIMA and GA-BiLSTM model. It is not difficult to find that the exponential smoothing model and ARIMA have a very obvious "lag" in places with large fluctuations, so manual adjustment is often required based on historical experience. Compared with the exponential smoothing model and ARIMA, GA-BiLSTM has obvious advantages.

5. Conclusion

Sales volume prediction of filling stations has received little attention in previous literature, but it is an important link. Meanwhile, it is also a difficult task due to large fluctuations, weak periodicity and being influenced by various objective factors. This paper proposes a combination model based on GA and BiLSTM. We add objective factors which influence the sales of filling stations to assist the prediction. One-hot encoding is used to encode the text variables and the high-dimensional text vectors are reduced by Embedding layer. Then, GA is used to adjust BiLSTM hyperparameters automatically. 940 days of sales data of a filling station in Kunming, China from 2019 to 2021 are taken as an example to discuss the effects of different feature combinations on prediction results. Comparisons have been made among the different ways of feature dimension reduction, different LSTM structures as well as

different GA parameters. The results show that: (1) feature dimension reduction using Embedding layer can effectively avoid model overfitting; (2) The accuracy of BiLSTM model is optimal; (3) BiLSTM hyperparameter optimization using GA can reduce the RMSE and MAE of the model by 3%. Finally, compared with other algorithms, the proposed model's test set accuracy can reach 89%, and its MAE and RMSE are better than other algorithms.

The proposed model does not take into account the impact of time steps on forecasting. To address this issue, attention mechanism can be added to the BiLSTM model, which can effectively improve the accuracy of multivariable prediction by screening out the information that is more important to the current task from a large amount of information.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (51874325) and Science Foundation of China University of Petroleum, Beijing (2462021BJRC009).

References

- Abdel-Aal, R.E., Al-Garni, A.Z., 1997. Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy* 22 (11), 1059–1069. [https://doi.org/10.1016/S0360-5442\(97\)00032-7](https://doi.org/10.1016/S0360-5442(97)00032-7).
- Asala Hl, Chebeir J, Zhu W, et al., A machine learning approach to optimize shale gas supply chain networks. *Proceedings - SPE Annual Technical Conference and Exhibition*. <https://doi.org/10.2118/187361-MS>.
- Asala, H.I., Chebeir, J.A., Manee, V., et al., 2019. In: An Integrated Machine-Learning Approach to Shale-Gas Supply-Chain Optimization and Refract Candidate Identification. *SPE Reservoir Evaluation and Engineering*, pp. 1201–1224. <https://doi.org/10.2118/187361-PA>.
- Bianco, V., Manca, O., Nardini, S., 2009. Electricity consumption forecasting in Italy using linear regression models. *Energy* 34 (9), 1413–1421. <https://doi.org/10.1016/j.energy.2009.06.034>.
- Chebeir, J., Asala, H., Manee, V., et al., 2019. Data driven techno-economic framework for the development of shale gas resources. *J. Nat. Gas Sci. Eng.* 72, 103007. <https://doi.org/10.1016/j.jngse.2019.103007>.
- Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- Ciulla, G., D'Amico, A., 2019. Building energy performance forecasting: a multiple linear regression approach. *Appl. Energy* 253, 113500. <https://doi.org/10.1016/j.apenergy.2019.113500>.
- Deb, C., Eang, L.S., Yang, J., et al., 2015. Forecasting energy consumption of institutional buildings in Singapore. *Procedia Eng.* 1734–1740. <https://doi.org/10.1016/j.proeng.2015.09.144>.
- Deb, C., Eang, L.S., Yang, J., et al., 2016. Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy Build.* 121, 284–297. <https://doi.org/10.1016/j.enbuild.2015.12.050>.
- Ediger, V.S., Akar, S., 2007. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Pol.* 35 (3), 1701–1708. <https://doi.org/10.1016/j.enpol.2006.05.009>.
- Efendi, R., Ismail, Z., Deris, M.M., 2015. A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand. *Appl. Soft Comput.* J. 28, 422–430. <https://doi.org/10.1016/j.asoc.2014.11.043>.
- Ekonomou, L., 2010 Feb. Greek long-term energy consumption prediction using artificial neural networks. *Energy* 35 (2), 512–517. <https://doi.org/10.1016/j.energy.2009.10.018>.
- Geurts, M., Box, G.E.P., Jenkins, G.M., 1977. Time series analysis: forecasting and control. *J. Market. Res.* 14 (2), 269. <https://doi.org/10.2307/3008255>.
- Gupta, V., Pandey, A., 2018. Crude oil price prediction using LSTM networks. *Int. J. Comput. Inf. Eng.* 12 (3), 226–230. <https://doi.org/10.5281/zenodo.1316680>.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Holland, J.H., 1975. In: *Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor University of Michigan Press. <https://doi.org/10.1086/418447>, 1975.
- Huang, L., Liao, Q., Qiu, R., et al., 2021a. Prediction-based analysis on power consumption gap under long-term emergency: a case in China under COVID-19. *Appl. Energy* 283, 116339. <https://doi.org/10.1016/j.apenergy.2020.116339>.
- Huang, L., Liao, Q., Zhang, H., et al., 2021b. Forecasting power consumption with an activation function combined grey model: a case study of China. *Int. J. Electr. Power Energy Syst.* 130, 106977. <https://doi.org/10.1016/j.ijepes.2021.106977>.
- Jiang, Z., Lin, B., 2012. China's energy demand and its characteristics in the industrialization and urbanization process. *Energy Pol.* 49, 608–615. <https://doi.org/10.1016/j.ijepes.2021.106977>.
- Laib, O., Khadir, M.T., Mihaylova, L., 2019. Toward efficient energy systems based on

- natural gas consumption prediction with LSTM Recurrent Neural Networks. *Energy* 177, 530–542. <https://doi.org/10.1016/j.energy.2019.04.075>.
- Li, W., Becker, D.M., 2021. Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy* 237, 121543. <https://doi.org/10.1016/j.energy.2021.121543>.
- Liang, Y., Li, M., Li, J., 2012. Hydraulic model optimization of a multi-product pipeline. *Petrol. Sci.* 9 (4), 521–526. <https://doi.org/10.1007/s12182-012-0237-2>.
- Lu, Q., Sun, S., Duan, H., et al., 2021. Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. *Energy Inf.* 4 (2), 1–20. <https://doi.org/10.1186/s42162-021-00166-4>.
- Ma, Z., Ye, C., Li, H., et al., 2018. Applying support vector machines to predict building energy consumption in China. *Energy Proc.* 780–786. <https://doi.org/10.1016/j.egypro.2018.09.245>.
- Sadaei, H.J., Guimarães, F.G., José da Silva, C., et al., 2017. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. *Int. J. Approx. Reason.* 83, 196–217. <https://doi.org/10.1016/j.ijar.2017.01.006>.
- Sagheer, A., Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* 323, 203–213. <https://doi.org/10.1016/j.neucom.2018.09.082>.
- Sahraei, M.A., Duman, H., Çodur, M.Y., et al., 2021. Prediction of transportation energy demand: multivariate adaptive regression splines. *Energy* 224, 120090. <https://doi.org/10.1016/j.energy.2021.120090>.
- Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45 (11), 2673–2681. <https://doi.org/10.1109/78.650093>.
- Siami-Namini, S., Tavakoli, N., Namin, A.S., 2019. The performance of LSTM and BiLSTM in forecasting time series. In: *Proceedings-2019 IEEE International Conference on Big Data*. Big Data, pp. 3285–3292. <https://doi.org/10.1109/BigData47090.2019.9005997>, 2019.
- Suganthi, L., Samuel, A.A., 2012. In: *Energy Models for Demand Forecasting - A Review*. Renewable and Sustainable Energy Reviews, pp. 1223–1240. <https://doi.org/10.1016/j.rser.2011.08.014>.
- Torrini, F.C., Souza, R.C., Cyrino Oliveira, F.L., et al., 2016. Long term electricity consumption forecast in Brazil: a fuzzy logic approach. *Soc. Econ. Plann. Sci.* 54, 18–27. <https://doi.org/10.1016/j.seps.2015.12.002>.
- Tulensalo, J., Seppänen, J., Ilin, A., 2020. An LSTM model for power grid loss prediction. *Elec. Power Syst. Res.* 189, 106823. <https://doi.org/10.1016/j.epsr.2020.106823>.
- Wang, B., Klemes, J.J., Zheng, T., et al., 2021. A fair profit allocation model for the distribution plan optimisation of refined products supply chains. *Comput. Aided Chem. Eng.* 1847–1852. <https://doi.org/10.1016/B978-0-323-88506-5.50286-2>.
- Wang, Q., Li, S., Li, R., 2018. Forecasting energy demand in China and India: using single-linear, hybrid-linear, and non-linear time series forecast techniques. *Energy* 161, 821–831. <https://doi.org/10.1016/j.energy.2018.07.168>.
- Wang, X., Zhan, H., Zhang, J., 2015. Research of oil product secondary distribution optimization based on collaborative distribution. *Procedia Comput. Sci.* 1367–1376. <https://doi.org/10.1016/j.procs.2015.08.210>.
- Wang, Y., Wang, J., Zhao, G., et al., 2012. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. *Energy Pol.* 48, 284–294. <https://doi.org/10.1016/j.enpol.2012.05.026>.
- Wei, N., Li, C., Peng, X., et al., 2019. Daily natural gas consumption forecasting via the application of a novel hybrid model. *Appl. Energy* 250, 358–368. <https://doi.org/10.1016/j.apenergy.2019.05.023>.
- Wei, X.T., Liao, Q., Zhang, H.R., et al., 2021. MILP formulations for highway petrol station replenishment in initiative distribution mode. *Petrol. Sci.* 18 (3), 994–1010. <https://doi.org/10.1007/s12182-021-00551-4>.
- Xu, L., Hou, L., Zhu, Z., et al., 2021. Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm. *Energy* 222, 119955. <https://doi.org/10.1016/j.energy.2021.119955>.
- Yuan, C., Liu, S., Fang, Z., 2016. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. *Energy* 100, 384–390. <https://doi.org/10.1016/j.energy.2016.02.001>.
- Zheng, J., Zhang, H., Dai, Y., et al., 2020. Time series prediction for output of multi-region solar power plants. *Appl. Energy* 257, 114001. <https://doi.org/10.1016/j.apenergy.2019.114001>.
- Zhu, L., Li, M.S., Wu, Q.H., et al., 2015. Short-term natural gas demand prediction based on support vector regression with false neighbours filtered. *Energy* 80, 428–436. <https://doi.org/10.1016/j.energy.2014.11.083>.