Computers & Industrial Engineering 176 (2023) 108935

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A comparative online sales forecasting analysis: Data mining techniques

Bo Zhang [a], Ming-Lang Tseng [b,c,d,e,*], Lili Qi [a], Yuehong Guo [a], Ching-Hsin Wang [b,f]

[a] School of Economics and Management, Tianjin University of Technology and Education, Tianjin, China
[b] Institute of Innovation and Circular Economy, Asia University, Taichung, Taiwan
[c] Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan
[d] UKM-Graduate School of Business, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
[e] Department of Business Administration, Asia University, Taichung, Taiwan
[f] Department of Leisure Industry Management, National Chin-Yi University of Technology, Taichung, Taiwan

ARTICLE INFO

Keywords:
Data mining
Online clothing sales
Sales forecasting
Optimization algorithm
Extreme learning machine

ABSTRACT

This study aims to improve the management efficiency of e-commerce platform and assists merchants on the e-commerce platforms in formulating a suitable sales plan urgently. Online sales forecasting analysis needs to be studied and shows that the management efficiency and operating income on an e-commerce platform is improved through accurate commodity sales forecasting. A novel online clothing sales forecasting model is proposed based on data mining technique. This study contributes to presenting the model references for e-commerce platform to make decisions on future sales and directions. (1) The gray correlation model was employed to mine the correlation degree between each feature and the clothing sales to select the features that have a great impact on clothing sales. (2) A sailfish optimization algorithm (SFO) algorithm with random disturbance strategy (SFOR) was proposed based on the SFO to improve the prediction effect of clothing sales. The benchmark function test results showed that the SFOR algorithm effectively avoided local extreme points. (3) The SFOR algorithm was used to solve the extreme learning machine (ELM) random parameter problem, and the SFOR-ELM-based online sales prediction model of clothing products suitable for multiple scenarios was constructed. In addition, three cases are applied to verify the SFOR-ELM-based online clothing sales forecast model. The verification results proved that SFOR-ELM achieved satisfactory prediction results, with its mean absolute percentage error values controlled below 5.1% and root mean square error values controlled below 16.2%.

1. Introduction

The global economy has entered the era of internet information techniques with the rapid popularization of internet information techniques, accelerating the transformation of the traditional economic development mode based on material production and material services to the new economic development mode based on information production and information services (Okpoti et al., 2019; Bag et al., 2021; Tseng et al., 2021). Driven by internet information techniques, e-commerce, as a brand-new business mode, has developed into a new economic growth point in countries all over the world (Boysen et al., 2019). Like offline sales, online sales based on e-commerce platforms are also highly competitive. Hence, the online platform needs to improve efficiency and provide customers with the most competitive price in a relatively short time to win customers in the current fierce competition, which requires the e-commerce platform to make an accurate prediction of future sales (Gautam and Singh, 2020; Tsoumakas, 2019; Zhang et al., 2021). The data mining technique must be applied in a mass online platform database.

At present, the literature has developed a variety of approaches to predict the sales of different commodities, including electricity, metals, clothing, agricultural products, etc. (Sano and Yamada, 2021; Giri et al., 2019). The forecasting methods are summarized into two categories: qualitative and quantitative forecasting (Marathe et al., 2019; Islam et al., 2020). The qualitative forecasting method is used to analyse various information obtained from market surveys, determine the price influencing factors, and make estimates and judgment on prices according to the development trend of the market. The advantages of qualitative forecasting methods include their simplicity and speed, but the disadvantage is that qualitative forecasting is affected by subjective factors and cannot provide an accurate numerical description of the development of price trends. Quantitative forecasting methods, which

* Corresponding author at: Institute of Innovation and Circular Economy, Asia University Taiwan, Taichung, Taiwan.
E-mail addresses: Tengminglang@gmail.com, Tsengminglang@asia.edu.tw (M.-L. Tseng).

https://doi.org/10.1016/j.cie.2022.108935
Received 10 November 2021; Received in revised form 4 September 2022; Accepted 20 December 2022
Available online 26 December 2022
0360-8352/© 2022 Elsevier Ltd. All rights reserved.

are not affected by subjective factors and focus on the analysis of the number of things in development, use statistical models to give a quantitative description of price trends based on historical statistical data (Choi and Lee, 2018).

In addition, quantitative forecasting methods are based on mathematical statistical models, deep learning algorithms, machine learning models and other methods to construct prediction models. Machine learning models have strong generalization and mapping capabilities. Several kinds of machine learning models have been devised, namely long-short term memory (LSTM) neural network, extreme learning machine, support vector machine, etc. (Lai et al., 2022; El Ouadi et al., 2022). The LSTM model is employed to predict commodity sales due to its powerful mapping ability. For instance, Weng et al. (2020) proposed a new supply chain sales forecasting method based on the LSTM model, which used LSTM to mine the time characteristics of supply chain sales data. However, Weng et al. (2020) ignored addressing the influence of random parameters on the forecast stability of the LSTM model. To this end, Shao et al. (2019) proposed an improved particle swarm optimization algorithm (IPSO) that was used to reduce the volatility of random parameters of the LSTM to improve the prediction performance, and then the optimized LSTM model was applied to predict nickel prices. The data mining technique is used for forecasting and improves the performance.

To accurately predict the sales of liquified natural gas (LNG) cylinders, Correia et al. (2020) constructed a combined model, which combined the time series model and the artificial neural network. The final prediction result was obtained through the Monte Carlo method. Additionally, Ji et al. (2019) combined the extreme gradient boosting (XGboost) and the autoregressive integrated moving average (ARIMA) to present a hybrid model to forecast the sales of cross-border e-commerce companies, using ARIMA to forecast the linear part and XGBoost to forecast the nonlinear part. Hybrid forecast models achieve a better fitting effect than single forecast models, however, Correia et al. (2020) and Ji et al. (2019) lacked to reduce the high computational cost of the combined model, which limits its application. To reduce the volatility of carbon prices, Sun et al. (2021) employed a decomposition-based approach to decompose the carbon price series, and the decomposed series were forecasted separately using the optimized neural network. The decomposition-based method reduces the volatility of the time series to a certain extent, but the complexity of the model is increased at the same time. However, Sun et al. (2021) ignored addressing the increased complexity of the decomposition-based method. In addition, Zhang et al. (2021) employed a two-layer decomposition method to decompose the crude oil price series, which greatly reduces the randomness of the crude oil price series. Meanwhile, a kernel extreme learning machine (KELM) was used to predict the decomposed crude oil price series, and a PSO-based optimization strategy was applied to improve the predictive accuracy of the KELM. However, Zhang et al. (2021) did not address the computational cost of two-order decomposition and the limitation of PSO algorithm which easily falls into local extremum. Karasu et al. (2020) developed a new crude oil price forecasting model using support vector regression (SVR), which combined feature extraction and multiobjective optimization techniques. However, Karasu et al. (2020) lacked to discuss the influence of super parameters on the forecast results of crude oil prices.

LSTM suffers from gradient disappearance in the training process, and the training time is too long because LSTM has a memory function. In addition, traditional single-hidden layer feed forward neural networks, such as artificial neural networks, use the gradient descent method to correct the threshold and weight of the network in the training process. The gradient descent method takes multiple iterations to correct the parameters, resulting in a long training time and sensitivity to the selection of the learning rate. The network may not converge when the learning rate is large; conversely, the convergence speed of the learning algorithm becomes slow and the training time of the network increases when the learning rate is small. In summary, to overcome the shortcomings of feedforward neural networks such as slow learning speed and long training time, Huang et al. (2006) proposed the extreme learning machine model (ELM). Due to its excellent predictive performance, the ELM model and its variants have been used in various fields (Silitonga et al., 2020; Tutuncu et al., 2021; Larrea et al., 2021). For instance, Kardani et al. (2021) combined the ELM model and artificial neural network model to forecast the permeability of tight carbonates, however, Kardani et al. (2021) failed to consider the disturbance of parameter randomness on the predictive effects. Therefore, Shariati et al. (2020) applied a novel heuristic algorithm to reduce the randomness of parameters of the ELM model to enhance the predictive stability, and employed the improved ELM model to forecast the compressive strength.

To improve the predictive accuracy, Tan and Zhang (2020) adopted a salp swarm algorithm-based optimization strategy to reduce the influence of parameter randomness. Although the model achieved accurate prediction, Tan and Zhang (2020) ignored that the salp swarm algorithm tends to fall into local solutions when solving high-dimensional problems. ELM models are used not only in the field of prediction, but also in classification. For instance, Yahia et al. (2020) presented a hybrid classification approach combining wave neural network and the ELM model, and proved the classification approach through simulation experiments. In addition, Albadr and Tiun (2020) combined ELM model with PSO-based optimization strategy for spoken language identification. In addition, ELM is widely used in the field of commodity price and sales forecasting. For instance, Chai et al. (2021) proposed an ELM-based carbon trading price prediction model, and used the PSO algorithm to optimize the model's parameters. To accurately predict stock market sales, Das et al. (2022) et al. proposed a combined model that integrates the improved crow search algorithm and ELM. Wu et al. (2021) used wave transform to reduce the noise signal in the stock price series, and used ELM to predict the denoised stock price. To improve the adaptability of the model to the nonlinearity of commodity price series, Mohanty et al. (2021) proposed a hybrid model that combines an autoencoder and ELM.

The input features have a direct impact on the prediction effect of the model. The stronger the correlation between input features and output features is, the higher the prediction accuracy of the model is. Common input feature selection methods include Pearson coefficient, correlation coefficient (Silva et al., 2022), gray correlation model (Wang et al., 2021) and so on. For instance, Chen et al. (2022) selected the input features of the temperature prediction model through the Pearson coefficient. In the process of building the battery capacity prediction model, Xu et al. (2021) determined the characteristics with the strongest correlation with the battery capacity through the gray correlation model. The gray correlation model is not sensitive to the number of samples and does not require the distribution of data. In addition, the gray correlation model has a small amount of calculation and a strong reliability. Therefore, gray correlation model has advantages in mining data information.

This study proposes a novel online sales forecast model of clothing products based on a data mining technique to realize the accurate prediction of multiscene and multicommodity sales. First, to mine the feature factors with the highest correlation degree with clothing sales, the grey correlation model was applied to calculate the correlation between clothing sales and each feature, and then the required feature factors were selected by the average correlation value. Second, an SFO-based improvement algorithm with stronger convergence performance was developed, called the SFOR algorithm. Several benchmark functions are employed to prove the convergence performance of the SFOR algorithm.

The statistical results show that the convergence performance of the SFOR algorithm outperforms the state-of-the-art heuristic algorithms. Additionally, the SFOR algorithm is employed to solve the problem that the random parameters disturb ELM forecasting results, and the SFOR-ELM is constructed for forecasting online sales of clothing products.

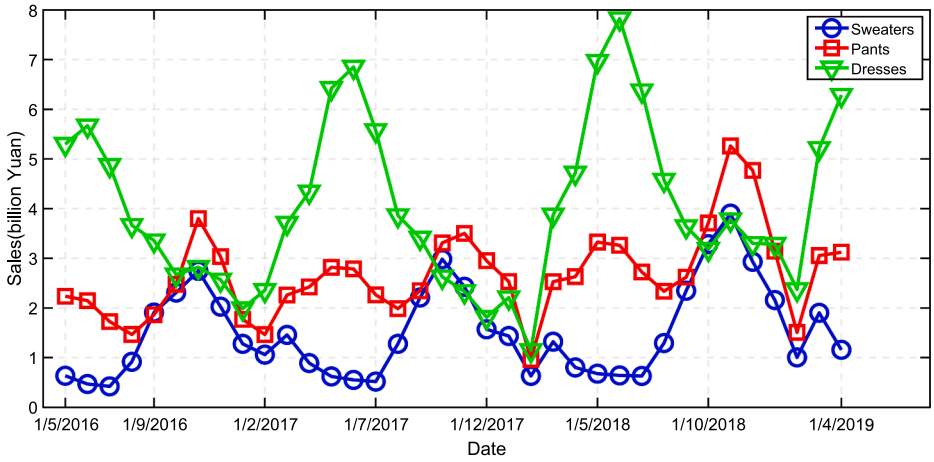

Fig. 1. Sales for three commodities.

Finally, three cases are applied to prove the SFOR-ELM model. For the three cases, the SFOR-ELM model obtains satisfactory results, which verifies the feasibility of the proposed approach. The contributions of this study are as follows:

- An SFOR algorithm with strong convergence performance is proposed and applied to the field of commodity sales forecasting.
- Combined with the data mining technique, the SFOR-ELM model is constructed to predict clothing sales.
- The obtained outcomes contribute to providing data and model references for e-commerce platforms to make decisions about future sales and improving economic efficiency.

The remaining chapters of this study are organized as follows. Section 2 mines and analyses the experimental data. Section 3 introduces the modelling approaches. Section 4 proves the proposed online clothing sales forecast method through three cases. Section 5 presents the contributions and limitations.

2. Acquisition and mining of experimental data

2.1. Experimental data acquisition

The statistical data set used in this study comes from an e-commerce platform in China, which counts the sales data of various clothing products from May 1, 2016 to April 1, 2019. The statistical data set includes sales and different features that can reflect sales. The detailed sale data of various clothing products are shown in appendix. Among them, ten features that best reflect sales (S; unit: billion Yuan) are counted: search popularity (X1; unit: million), search heat (X2; unit: million), number of visitors (X3; unit: 100 million), pageviews (X4; unit: billion), number of favorites (X5; unit: 10 million), collection times (X6; unit: 10 million), number of additional buyers (X7; unit: 10 million), number of additional purchases (X8; unit: 10 million), customer group index (X9; unit: 100 thousand), transaction index (X10; unit: million). Features X1 and X2 refer to the number of search times of the commodity. The higher the values of X1 and X2 are, the greater the likelihood that the commodity is purchased. Features X3 and X4 reflect the number of times of the commodity website has been visited or viewed. However, there is a difference in definition between X3 and X4. X3 refers to the number of visitors to a single IP. X4 is the number of times each page of the store was viewed. The indicator value of X4 is accumulated when a user opens or refreshes the same page multiple times. The higher values of the X5 and X6 are, the higher the popularity and traffic of the store is, and the higher the sales are. Features X7 and X8 reflect the number of times the commodity has been placed in the shopping cart. The higher values of X7 and X8 are, the more likely the commodity is purchased. Feature X9 helps stores locate the gap between commodities and the market, and analyzes the pros and cons of different industries. Generally, the higher the X9 is, the higher the number of buyers is. Feature X10 reflects the popularity of the commodity. Generally, the higher the X10 is, the higher the sales are.

Additionally, this study counted the sales data of a variety of clothing products on the e-commerce platform, such as sweaters, pants, dresses, down jackets, shirts, etc. This study selected representative sweaters, trousers, and dresses as the research objects, and analyzed the influence of different features on clothing sales. The sales curves of sweaters, pants, and dresses from May 1, 2016 to April 1, 2019 are depicted in Fig. 1.

Fig. 1 showed that the sales curve of the three clothing commodities shows a certain periodicity. For dresses, sales peak in May and June every year. For pants, sales peak in April or November each year. For sweaters, sales peak in November each year. To consider the periodic characteristics of clothing commodity sales, the annual commodity sales are predicted when building the forecasting model.

2.2. Experimental data mining

Data mining is based on interdisciplinary intersection and fusion, covering intelligent algorithms, machine learning models, mathematical statistics and other fields. The intrinsic laws in the data set can be mined through the analysis of massive data, fuzzy data and random data. Common data mining types include cluster analysis, classification and regression, correlation rule analysis and time series analysis. Cluster analysis is to gather similar data together to achieve the purpose of classification. Classification and regression are based on predictive approaches to model unknown data. The obtained results of classification are discrete, while the results of regression are continuous. Correlation rule analysis is to study the correlation degree between items in data set. Time series analysis aims to study the relationship between time series. This research is mainly based on classification and regression in data mining, and correlation rule analysis to study the correlation degree each feature variable to sales, and to predict the sales of different commodities.

Common feature extraction methods are mutual information and fisher score, furthermore, gray correlation model are used for input feature selection. In addition, gray correlation model has the advantages of low sensitivity to sample size and low requirement for regularity of data. Therefore, this study employed the gray correlation model to quantitatively analyze the influence of different features on sales.

The dimensions of features and sales are different, and the quantity values differ greatly. Therefore, the data needs to be processed to

Table 1
The results of grey correlation calculation.

r_0										
Features	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Sweaters	0.73	0.79	0.70	0.86	0.74	0.79	0.79	0.90	0.72	0.77
Pants	0.73	0.78	0.67	0.84	0.67	0.72	0.74	0.90	0.78	0.86
Dresses	0.74	0.70	0.65	0.84	0.70	0.76	0.72	0.90	0.81	0.82
Average G_c	0.73	0.76	0.67	**0.85**	0.70	0.76	0.75	**0.90**	**0.77**	**0.82**

eliminate the differences between the data. At present, different dimensionless processing methods have been developed, such as initial value method, the extreme value method and the mean value method. In this study, the initial value method is applied to eliminate the differences between the data, that is, the first value in the same sequence is divided by the subsequent data, and the values of the new sequence obtained are all bigger than 0. The parent sequence after data transformation is denoted as $\{S_0\}$, and the subsequence is $\{S_i\}$, then the correlation coefficient C^t_0 between the parent sequence and the subsequence at time t is calculated as follows:

$$\begin{cases} C_0^t = \dfrac{\Delta\min + \nu\Delta\max}{\Delta_0^t + \nu\Delta\max} \\ \Delta_0^t = \left|s_0^t - s_i^t\right|, \ (1 \leqslant i \leqslant m) \end{cases} \tag{1}$$

where $\triangle^t_0$ represents the absolute difference between the parent sequence and the subsequence at time t; $\triangle\max$ is the maximum value of the absolute difference; $\nu(\nu = 0.1)$ is the resolution coefficient to avoid distortion caused by excessive difference; $\triangle\min$ is the minimum value of the absolute difference; m is the number of subsequences.

Then the correlation degree r_0 between the parent sequence S_0 and the subsequence S_i is obtained:

$$r_0 = \frac{1}{Nm}\sum_{i=1}^{Nm} C_{0i}^t \tag{2}$$

where Nm is the number of data in the sequence.

Finally, the correlation sequence of m subsequences to the same parent sequence can be obtained. Based on the gray correlation model, the calculated correlation degree is listed in Table 1.

The selection of model input features is directly related to the prediction effect of the model. The prediction effect of the model will not be improved if the correlation between the selected input features and output features is poor, and the calculation cost of the model is increased. Through the gray correlation model, the characteristic factors with strong correlation with commodity sales are selected and used as the input characteristics of the prediction model, which can not only reduce the calculation cost of the model and reduce the difficulty of model regression, but also improve the prediction performance of the model.

Table 1 showed the correlation degree values between different features and clothing sales. The results revealed that the correlation degree between X8 and sales was the highest. For three different clothing products, the correlation degree values between X8 and sales reached 0.9. Additionally, features and sales exhibited a strong correlation, the values of G_C of X4 and X10 reached more than 0.8. The values of G_C of X1-X2 and X5-X7 all reached 0.7. However, feature X3 exhibited the lowest correlation with clothing sales, and the average value of r_0 was smaller than 0.7. The information contained in X3 and X4 is different. The number of visitors refers to recording the number of unique visitors who visit the website in one day. If the same visitors visit the same website in one day, the number of unique visitors will not increase cumulatively. Pageviews refers to the number of times a visitor visits a page of a website in a day. If the same visitor visits the same page of the website multiple times in one day, the number of page views will increase cumulatively. When consumers browse the same website repeatedly, it indicates that consumers have a higher desire to buy this product. Therefore, X4 has the highest correlation with commodity sales compared with X3. To simplify the calculation of the prediction model, the four features most related to sales are selected as the input features of the model. Fig. 2 presents the relationship curve between features X4, X8, X9, X10 and clothing sales.

Fig. 2 revealed that the fluctuation trend of the curves of features X4, X8, X9, X10 reflects the fluctuation trend of the clothing sales curve. Among them, the feature X8 curve has the highest fitting degree to the sales curve, and the fluctuation of the feature X8 curve is basically consistent with the fluctuation of the sales curve.

The input features of the predictive model have a direct impact on the predictive effects. If the correlation degree between the input features and the output features is not enough, the predictive effect of the model will deteriorate; if the number of input features is too large, not only will it not improve the predictive effect, but also it will increase the amount of calculation of the model. It is necessary to mine the impact of different features on sales, quantitatively analyze the degree of correlation between input features and output features, and select the input features with the highest correlation degree with output features. The analysis results of the correlation model in Table 1 proved that the features X4, X8, X9, and X10 had the highest correlation degree with clothing sales. Hence, this study selected the features X4, X8, X9, and X10 as the input features of the clothing sales forecasting model.

3. Model construction

The online clothing sales forecasting process based on data mining technique is depicted in Fig. 3.

3.1. Extreme learning machine

Huang et al. (2006) proposed an advanced feedforward neural network model, namely extreme learning machine. The learning efficiency of the ELM model is higher compared with the traditional neural network model. In addition, the weights and thresholds do not need to be updated during the model training process, which makes the ELM model has the characteristics of fast training speed, simple structure, and strong generalization ability. On the basis of the above advantages, ELM model is widely used in various fields, such as fault classification, image recognition, life prediction, etc. The structure of the ELM model is presented in Fig. 4.

Given M random samples $\{x_i, t_i\}$, where the input vector $x_i = [x_{i1}, x_{i2}, \ldots, x_{iM}]^T \in R^n$ and the output vector $t_i = [t_{i1}, t_{i2}, \ldots, t_{iM}]^T \in R^m$, then the output function of the feedforward neural network is obtained as follows (Silitonga et al., 2020):

$$\sum_{i=1}^{N_h} \rho_i g(w_i \cdot x_j + \beta_i) = o_j \ (j = 1, 2, \ldots M) \tag{3}$$

where N_h is the number of hidden layer neurons; β_i is the bias value of the i-th hidden layer neuron; o is the output vector of the feedforward neural network; g is the activation function; $\rho_i = [\rho_{i1}, \rho_{i2}, \ldots, \rho_{im}]^T$ is the weight vector connecting the output layer neuron and the hidden layer neuron; $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector connecting the input layer neuron and the hidden layer neuron.

The weight vector and threshold vector are randomly initialized

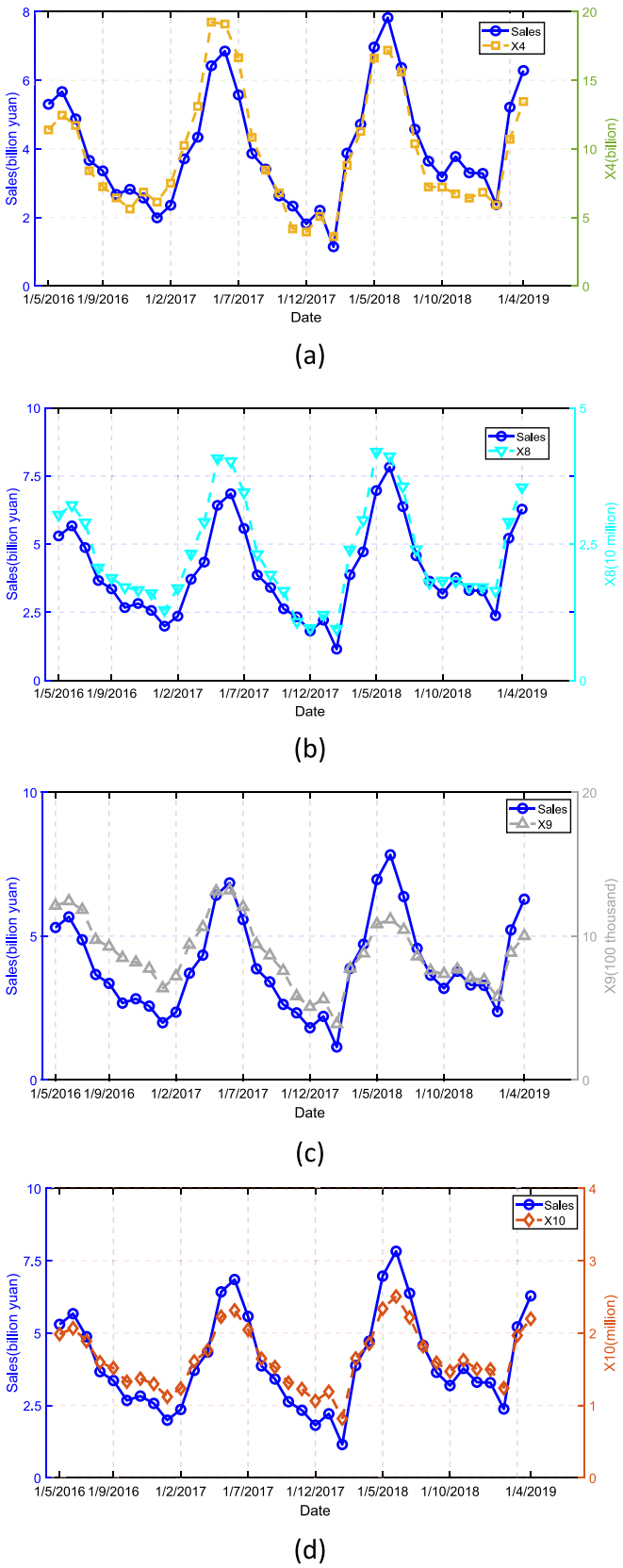

Fig. 2. Feature variables and sales.

when the number of hidden layer neurons is equal to the number of training set, and the feedforward neural network can approximate the training output vector with 0 error, namely:

$$\sum_{i=1}^{N_h} \|o_j - t_j\| = 0 \ (j = 1, 2, \ldots M) \tag{4}$$

Thus, the ELM output function of M random samples is obtained:

$$\sum_{i=1}^{N_h} \rho_i g(w_i \cdot x_j + \beta_i) = t_j \ (j = 1, 2, \ldots M) \tag{5}$$

The matrix form of Equation (5) is as follows:

$$H\rho = T \tag{6}$$

where H is the hidden layer output matrix as follows:

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + \beta_1) & g(w_2 \cdot x_1 + \beta_2) & \cdots & g(w_l \cdot x_1 + \beta_l) \\ g(w_1 \cdot x_2 + \beta_1) & g(w_2 \cdot x_2 + \beta_2) & \cdots & g(w_l \cdot x_2 + \beta_l) \\ \vdots & \vdots & \vdots & \vdots \\ g(w_1 \cdot x_M + \beta_1) & g(w_2 \cdot x_M + \beta_2) & \cdots & g(w_l \cdot x_M + \beta_l) \end{bmatrix} \tag{7}$$

where l is the number of neurons in the hidden layer.

The weight vector of output layer and hidden layer is obtained by the least square solution of the equations (Tutuncu et al., 2021):

$$\min_{\rho} \|H\rho - T\| \tag{8}$$

The solution of Equation (8) is as follows:

$$\widehat{\rho} = H^{+}T \tag{9}$$

where H^{+} is the generalized inverse of H.

ELM model does not need to update the weight vector and the threshold vector in the iterative process, which speeds up the training of the ELM model. However, the random initialization of the weight vector and the threshold vector has an impact on the predictive effects of the ELM model. The number of nodes in the hidden layer of the ELM will increase and the prediction performance will decrease when the random weight vector and the random threshold vector are set incorrectly. Therefore, how to select the appropriate weight vector and threshold vector of ELM so as to give play to the best prediction performance is the key to achieving good predictive results. At present, most researches employ *meta*-heuristic algorithm to optimize ELM model's random parameter vector. However, the convergence performance of different heuristic algorithms is different, so developing a novel heuristic algorithm suitable for forecasting the online clothing sales is necessary. This research uses SFOR algorithm to solve this limitation of ELM model.

3.2. SFOR algorithm

Sailfish is the fastest fish in the ocean, and its maximum speed can reach 100 km/h. In the process of foraging, the individuals of the sailfish school cooperate with each other to achieve the purpose of prey hunting by surrounding, hunting, and attacking the prey. For example, the sailfish school uses rostrum to alternately attack the sardine school when the sailfish school is preying on sardines. During this period, the sardines are constantly injured. The injured sardines will leave the fish school and be preyed by the sailfishes. Shadravan et al. (2019) proposed the sailfish optimization algorithm (SFO) by modeling the predation behavior of sailfish school and the escape behavior of sardine school. SFO algorithm includes initialization, elite strategy, alternate attack strategy, and predation strategy.

In the SFO algorithm, the sailfish school represents the candidate solution set, and the variable of the problem represents the search space. The SFO algorithm can address optimization problems of different dimensions, such as one-dimensional, two-dimensional, and high-dimensional problems. Assuming that the search space of sailfish school is *Dim* dimension, the position of the *i*-th sailfish when searching

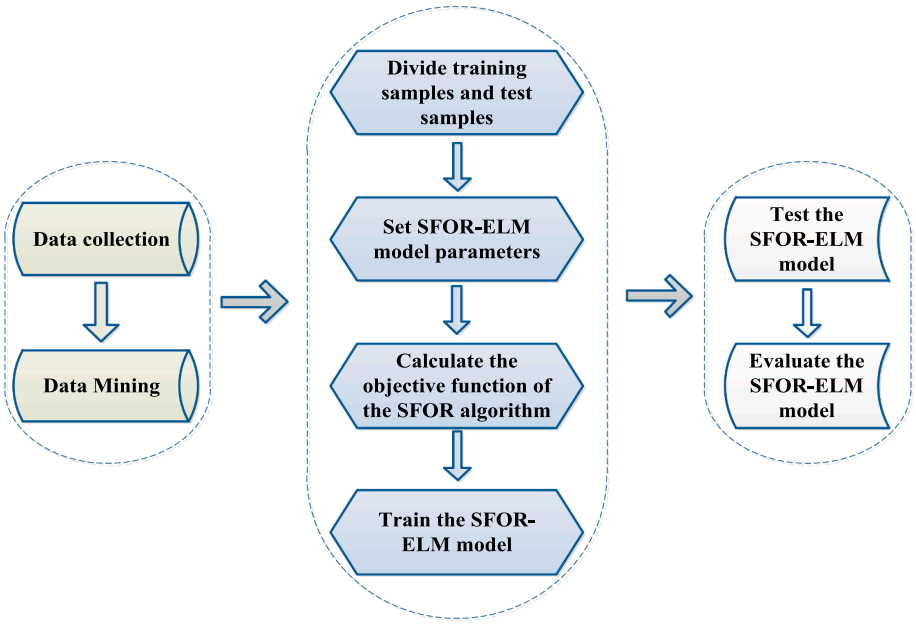

Fig. 3. Flow chart of clothing sales forecast.

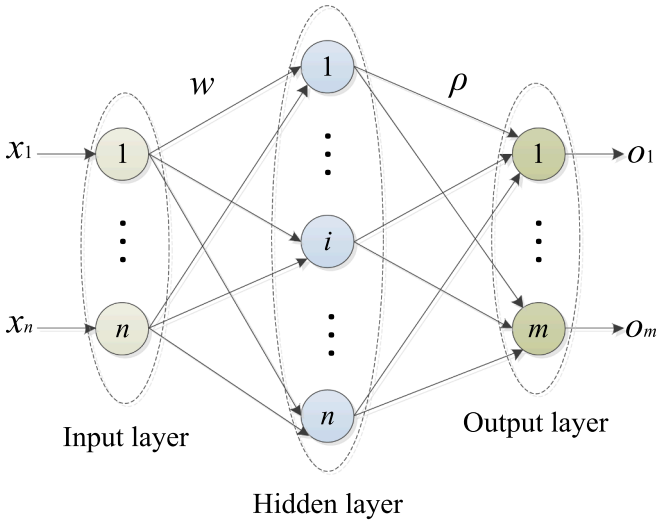

Fig. 4. ELM model topology.

in the j-th dimension space is $SF_{i,j}(i = 1,2,\ldots,N_{SF})$, where N_{SF} represents the number of sailfishes. Therefore, the position matrix of the sailfish school is expressed as follows (Nassef et al., 2021):

$$SF = \begin{bmatrix} SF_{1,1} & SF_{1,2} & \cdots & SF_{1,Dim} \\ SF_{2,1} & SF_{2,2} & \cdots & SF_{2,Dim} \\ \vdots & \vdots & \vdots & \vdots \\ SF_{N_{SF},1} & SF_{N_{SF},2} & \cdots & SF_{N_{SF},Dim} \end{bmatrix} \tag{10}$$

where Dim represents the number of variables to be solved.

The fitness value of the sailfish school is calculated using the fitness function as follows:

$$Fit(SF) = Fit(SF_1, SF_2, \ldots, SF_{N_{SF}}) \tag{11}$$

where SF_i represents the position matrix of the i-th sailfish.

Then the fitness value matrix SF_{Fit} corresponding to the position matrix of sailfish school is obtained as follows (El Hammouti et al., 2019):

$$SF_{Fit} = \begin{bmatrix} Fit(SF_{1,1}, SF_{1,2}, \ldots, SF_{1,Dim}) \\ Fit(SF_{2,1}, SF_{2,2}, \ldots, SF_{2,Dim}) \\ \vdots \\ Fit(SF_{N_{SF},1}, SF_{N_{SF},2}, \ldots, SF_{N_{SF},Dim}) \end{bmatrix} = \begin{bmatrix} Fit_{SF_1} \\ Fit_{SF_2} \\ \vdots \\ Fit_{SF_{N_{SF}}} \end{bmatrix} \tag{12}$$

where Fit_{SFi} represents the target value of the i-th sailfish.

The sardine school in the SFO algorithm is another important candidate solution set, which converges to the optimal solution through the cooperation of the sardines and the sailfishes. The sailfish school is distributed in the search space and is the key search force in the optimization process of the SFO algorithm, meanwhile, the sardine school assists the sailfish school in searching for the optimal solution. Assuming that sardine school is also searching in the Dim-dimensional space, the position of the i-th sardine in the j-dimensional space is expressed as $SD_{i,j}$. The position matrix SD of the sardine school is as follows:

$$SD = \begin{bmatrix} SD_{1,1} & SD_{1,2} & \cdots & SD_{1,Dim} \\ SD_{2,1} & SD_{2,2} & \cdots & SD_{2,Dim} \\ \vdots & \vdots & \vdots & \vdots \\ SD_{N_{SF},1} & SD_{N_{SF},2} & \cdots & SD_{N_{SF},Dim} \end{bmatrix} \tag{13}$$

where N_{SD} represents the number of sardines, and the position matrix SD is used to store the position of each sardine.

Similarly, the fitness value matrix SD_{Fit} is obtained through calculating the fitness value of sardine school:

$$SD_{Fit} = \begin{bmatrix} Fit(SD_{1,1}, SD_{1,2}, \ldots, SD_{1,Dim}) \\ Fit(SD_{2,1}, SD_{2,2}, \ldots, SD_{2,Dim}) \\ \vdots \\ Fit(SD_{N_{SD},1}, SD_{N_{SD},2}, \ldots, SD_{N_{SD},Dim}) \end{bmatrix} = \begin{bmatrix} Fit_{SD_1} \\ Fit_{SD_2} \\ \vdots \\ Fit_{SD_{N_{SD}}} \end{bmatrix} \tag{14}$$

where Fit_{SDi} represents the fitness value of the i-th sardine.

The elite selection strategy in the SFO algorithm is used to avoid losing the best solution, and the best sailfish position is used as an elite solution in the iterative process of the SFO algorithm. During the t_{th} iteration, the elite sailfish SF^t_{elite} affects the maneuverability of the sardines during the predation process, thereby catching the sardines. In addition, the sardines will be injured during the hunting process of the sailfish school, and the position of the injured sardine is selected as the best target position. Therefore, in the elite strategy, the elite sailfish SF^t_{elite} and the injured sardine SD^t_{injury} are selected as two elite solutions to update the positions of the remaining sailfishes and sardines (Wankhede et al., 2020).

The sailfish suddenly attacked during the hunting process, and then adopted the strategy of alternate attack. Meanwhile, the sailfish update its position based on partner's position, elite swordfish and injured sardine. At this time, the position of sailfish is as follows:

$$P^t_{SF} = P^t_{elite} - \theta_t \times \left(\left(\frac{SF^t_{elite} + SD^t_{injury}}{2}\right) \times rand - P^t_{SF_{current}}\right) \quad (15)$$

where $P^t_{SFcurrent}$ represents the current position of the sailfish and *rand* represents the random number between 0 and 1. θ_t is calculated as follows:

$$\theta_t = (2 \times rand - 1) \times d_{SD} \quad (16)$$

where d_{SD} represents sardine density.

The number of sardines and the density decrease during the predation process of the sailfish. The parameter d_{SD} is calculated according to the number of sardines and sailfishes.

$$d_{SD} = \frac{N_{SD}}{N_{SD} + N_{SF}} \quad (17)$$

The energy of the sardines is continuously reduced when the sailfishes prey on the sardines, which weakens the sardines' mobility and direction discrimination ability, thus causing the sardines to escape from the group and be preyed by the sailfishes. During the escape process of the sardines, the position of the sardine P^t_{SD} is updated according to the position of the elite sailfish and the current position, as follows:

$$P^t_{SD} = rand \times (P^t_{elite} - P^t_{SD_{current}} + A_P) \quad (18)$$

$$A_P = A \times (1 - 2 \times t \times e) \quad (19)$$

where $P^t_{SDcurrent}$ represents the current position of the sailfish; the coefficient A_p is used to represent the attack power of the sailfish; and t is the current iteration number. The coefficients A and e are employed to simulate the process of A_p linearly decreasing.

Additionally, parameters are used to determine the number of sardines that need to be replaced during the iteration processes of the SFO algorithm. The number of replaced sardines N_a and the number of variables N_b are calculated as follows:

$$\begin{cases} N_a = A_p \times N_{SD} \\ N_b = A_p \times d_t \end{cases} \quad (20)$$

where d_t represents the number of variables in the t-th iteration.

The attack power of the sailfish is relatively weak when $Ap < 0.5$, only the N_b variable of N_a sardines needs to be updated with the new position; the attack power of the sailfish is relatively strong when $Ap \geq 0.5$, at this time, all the positions of sardines need to be updated.

Finally, the position of the sailfish is replaced with the position of the sardine when the fitness value of the sardine outperforms the fitness value of the sailfish, as follows:

$$P^t_{NF} = P^t_{ND} \ (if \ Fit_{SDt} < Fit_{SFt}) \quad (21)$$

where Fit_{SDt} and Fit_{SFt} respectively represent the fitness values of sardines and sailfish at the t-th iteration.

Shadravan et al. (2019) compared the SFO algorithm with the state-of-the-art heuristic algorithms and proved the superiority of the SFO algorithm. However, the convergence result of the SFO algorithm has not reached the best state in the validation process of using the benchmark functions, and its convergence ability needs to be further improved.

The positions of sailfishes and sardines in the SFO algorithm represent candidate solutions. The convergence range of the SFO algorithm becomes smaller in the later iteration, and the diversity of the solution becomes worse. This makes the algorithm's ability to jump out of local extremes worse, causing the population to fail to converge to global optimal solution. To solve this limitation, this study proposed the SFO algorithm based on random disturbance strategy. In the iteration processes of the SFOR algorithm, the positions of the sailfishes and sardines are disturbed to improve the ability of the SFO algorithm to get rid of local extremes. The specific steps are as follows:

Table 2
Pseudocode of the SFOR algorithm.

Initialize the positions of the sailfish school and the sardine school in the SFOR algorithm using equations (10) and (13).
Calculate the fitness values of sailfish and sardines using equations (12) and (14).
Select SF^t_{elite} and SD^t_{injury}.
Start iteration
for 1: The maximum number of iterations
if $d_c < 0.4$
for 1:N_{SF}
Calculate θ_t using Equation (16).
Update sailfish position by Equation (15).
end
Calculate A_p using Equation (19).
If $A_p < 0.5$
Calculate N_a and N_b using Equation (20).
Update sardines position based on N_a and N_b using Equation (23).
else
Update the positions of all the sardines using Equation (23).
end
else
Update the positions of sailfish and sardines using Equation (22).
Calculate the fitness values of sailfish and sardines.
Compare the best positions of sailfish and sardines using Equation (21).
end
end
Output the optimal solution of the problem to be optimized.

(1) Set the disturbance coefficient d_c that is a random number between 0 and 1;
(2) Determine whether to disturb the positions of sailfishes and sardines;
(3) Update the positions of sailfishes and sardines when $d_c < 0.4$ using Eqs. (15) and (18);
(4) Disturb the positions of sailfishes and sardines when $d_c \geq 0.4$ using Equation (22), as follows:

$$\begin{cases} P^t_{SF} = P^t_{SFcurrent} + rand \times (P_{best} - P^t_{SF_{current}}) \times \exp(Fit(P_{best}) - Fit(P^t_{SF_{current}})) \\ P^t_{SD} = P^t_{SDcurrent} + rand \times (P_{best} - P^t_{SD_{current}}) \times \exp(Fit(P_{best}) - Fit(P^t_{SD_{current}})) \end{cases} \quad (22)$$

where P_{best} represents the global best position, $Fit(P_{best})$ represents the fitness value of P_{best}, $Fit(P^t_{SFcurrent})$ represents the current fitness value of the sailfish, and $Fit(P^t_{SDcurrent})$ represents the current fitness value of the sardine.

The disturbance frequency is related to the threshold of the disturbance coefficient. If the threshold is set too large, the influence of the disturbance coefficient on the algorithm will be small, resulting in poor diversity of the population; if the threshold is set too small, the computational cost of the model will increase. To balance the global and local convergence ability of the SFOR algorithm, the threshold of disturbance coefficient is set to 0.4 in this study.

(5) Calculate the fitness values of sailfishes and sardines.

The problem that the population diversity becomes worse in the later stage of the SFO algorithm is addressed through the random disturbance strategy. In addition, the global optimization ability and the ability to jump out of local extreme solutions are enhanced.

The injured sardine SD^t_{injury} as an elite solution in the SFO algorithm has a greater impact on the convergence performance. However, the influence of the current position of the elite sailfish and sardine is considered only in Equation (18). The influence of the elite solution SD^t_{injury} is not considered, which affects the convergence performance of the SFO algorithm. Therefore, this research has improved the location

Table 3
Unimodal benchmark functions.

Benchmark functions	Optima	Range	*Dim*
$f_1(u) = \sum_{j=1}^{n} u_j^2$	0	[-100, 100]	30
$f_2(u) = \sum_{j=1}^{n} \lvert u_j \rvert + \prod_{j=1}^{n} \lvert u_j \rvert$	0	[-10, 10]	30
$f_3(u) = \max \lvert u_j \rvert \ (1 \leqslant j \leqslant u)$	0	[-100, 100]	30
$f_4(u) = \sum_{j=1}^{n-1} [(u_j - 1)^2 + 100(u_{j+1} - u_j^2)]$	0	[-30, 30]	30

update strategy of sardines to further explore the convergence ability of the SFO algorithm, as follows:

$$P_{SD}^{t} = rand \times \left(\frac{SF_{elite}^{t} + SD_{injury}^{t}}{2} \times r - P_{SD_{current}}^{t} + A_P\right) \tag{23}$$

where r represents a random number between 0 and 1.
The pseudo code of the SFOR algorithm is listed in Table 2.

3.3. SFOR-ELM model

The analysis in Section 3.1 presented that the predictive effect of the ELM model was affected by the random weight vector and the random threshold vector. Therefore, a novel online sales forecast approach of clothing products was constructed using the ELM optimized by the SFOR-based strategy (SFOR-ELM). Data mining technique in Section 2 was applied to select the four features that affected the clothing sales, and selected four features were taken as the input features of the SFOR-ELM model. Online sales forecast of clothing products is summarized as the following steps:

(1) Count data set from e-commerce platform. The data set contains clothing sales and different feature factors.
(2) Employ data mining technique to determine the features that have the greatest impact on clothing sales, and take the selected features as the input features of the SFOR-ELM model.
(3) Divide the data set into the training set and test set of the SFOR-ELM model.
(4) Set the parameters of the SFOR-ELM model.
(5) Take mean square error (MSE) as the objective function of the SFOR algorithm:

$$MSE = \frac{1}{ns} \times \sum_{i=1}^{ns} (p_i - a_i)^2 \tag{24}$$

where p_i is the predicted value of clothing sales; a_i is the actual value of clothing sales; and ns is the number of samples.

(6) Use the training set to train the online clothing sales forecast model.
(7) Use the testing set to test the online clothing sales forecast model.
(8) Analyze the predictive performance of the SFOR-ELM model using evaluation indicators. This study employed three evaluation indicators to evaluate the prediction results, as follows:

$$MAPE = \frac{1}{ns} \times \sum_{i=1}^{ns} \left| \frac{p_i - a_i}{a_i} \right| \times 100\% \tag{25}$$

$$RMSE = \sqrt{\frac{1}{ns} \times \sum_{i=1}^{ns} (p_i - a_i)^2} \tag{26}$$

$$R2 = 1 - \frac{\sum_{i=1}^{ns} (p_i - a_i)^2}{\sum_{i=1}^{ns} (a_i - p_m)^2} \tag{27}$$

where p_m is the predicted mean value of online clothing sales.

4. Case studies

4.1. SFOR algorithm convergence performance test and significance test

A set of benchmark functions that include unimodal and multimodal benchmark functions are applied to prove the convergence performance of the SFOR algorithm, as listed in Tables 3 and 4 (Liu et al., 2021; Li et al., 2021).

Table 3 presented four unimodal benchmark functions. *Dim* was 30, and the optimal value was 0. The unimodal benchmark functions do not have local extreme points, and only have a global optimal solution, which is suitable for testing the global convergence ability of the optimization algorithms. Table 4 showed three multimodal benchmark functions. The multimodal benchmark functions have more local extreme points compared with the unimodal benchmark functions, which brings challenges to the optimization algorithms. Meanwhile, multimodal benchmark functions are suitable for testing the ability of the optimization algorithm to jump out of the local extremum.

In addition, to test the convergence effect of the proposed SFOR algorithm, the salp swarm algorithm (SSA) (Mirjalili et al., 2017), SFO, crow search algorithm (CSA) (Askarzadeh, 2016), multi-verse optimizer (MVO) (Mirjalili et al., 2016) proposed in recent years were selected as the comparison algorithms. The selected comparison algorithms are representative and have been applied in various fields, such as parameter optimization and pattern recognition.

To make the statistical results more objective, a unified platform is used to test the algorithms, and each algorithm is run 30 times for each benchmark function. The optimal value, worst value, average value, standard deviation (*std*), average value, and average running time (T) of the 30 running results are counted. The higher the complexity of the algorithm is, the longer the running time is. To this end, the complexity of the SFOR algorithm is analyzed by comparing the average running time of each algorithm. For the selected optimization algorithms, the maximum number of iterations and the initial population size are set to 1000 and 30, respectively, and the remaining adjustable parameters of the algorithms maintain the initial values. Tables 5 and 6 list the statistical results.

Table 5 and Table 6 presented the statistical values of the unimodal and multimodal benchmark functions. Table 5 revealed that the SFOR algorithm obtained more satisfactory statistical results than the state-of-the-art algorithms for the unimodal benchmark functions. For f_1-f_4, the SFOR algorithm exhibited strong convergence ability, and the convergence values were closest to the global optimal value 0. The position update strategy in the SFOR algorithm of the sardine considers not only the influence of the elite sailfish, but also the influence of the injured sardine, which strengthens the search ability of the sardine.

For multimodal benchmark functions, the statistical results of the

Table 4
Multimodal benchmark functions.

Benchmark functions	Optimal solution	Range	*Dim*
$f_5(u) = \sum_{j=1}^{n} [u_j^2 - 10\cos(2\pi u_j) + 10]$	0	[-5.12, 5.12]	30
$f_6(u) = -20\sum_{j=1}^{n} \exp(-0.2\sqrt{\frac{1}{n}\sum_{j=1}^{n} u_j^2}) - \exp(\frac{1}{n}\sum_{j=1}^{n} \cos(2\pi u_j)) + 20 + e$	0	[–32, 32]	30
$f_7(u) = \frac{1}{4000}\sum_{j=1}^{n} u_j^2 - \prod_{j=1}^{n} \cos(\frac{u_j}{\sqrt{j}}) + 1$	0	[-600, 600]	30

Table 5
The statistical results for unimodal benchmark functions.

f	Algorithm	Statistical results				
		Optimal convergence	Worst convergence	Mean convergence	*std*	T(s)
f_1	SSA	8.71E-09	1.81E-08	1.16E-08	2.22E-09	0.30
	CSA	3.45E-02	3.09E-01	8.80E-02	5.23E-02	0.12
	MVO	1.42E-01	5.11E-01	3.25E-01	1.09E-01	0.51
	SFO	2.16E-13	7.07E-10	1.41E-10	2.11E-10	1.01
	SFOR	**8.89E-128**	**6.55E-64**	**2.19E-65**	**1.19E-64**	0.18
f_2	SSA	6.12E-01	4.27	1.27	1.21	0.32
	CSA	6.90E-01	4.30	2.05	0.97	0.14
	MVO	2.21E-01	8.36E-01	4.42E-01	1.47E-01	0.41
	SFO	6.07E-06	8.59E-05	3.36E-05	2.09E-05	1.01
	SFOR	**9.62E-65**	**1.26E-31**	**2.31E-32**	**4.22E-33**	0.24
f_3	SSA	1.61	13.69	8.25	3.09	0.27
	CSA	2.42	6.53	4.22	1.06	0.11
	MVO	3.55E-01	1.65	9.13E-01	2.80E-01	0.43
	SFO	6.09E-08	8.52E-06	1.65E-06	1.23E-06	0.97
	SFOR	**4.98E-62**	**1.91E-32**	**9.90E-34**	**3.63E-33**	0.18
f_4	SSA	9.45	327.97	61.25	76.96	0.34
	CSA	25.98	610.45	106.23	119.59	0.14
	MVO	29.48	2.61E + 03	383.92	683.49	0.46
	SFO	4.50E-04	1.00E-01	2.58E-02	2.91E-02	1.01
	SFOR	**7.53E-05**	**4.89E-01**	**1.47E-01**	**1.36E-01**	0.26

Table 6
The statistical results for multimodal benchmark functions.

f	Algorithm	Statistical results				
		Optimal convergence	Worst convergence	Mean convergence	*std*	T(s)
f_5	SSA	18.90	114.41	59.53	20.95	0.29
	CSA	13.94	47.05	27.50	9.09	0.13
	MVO	54.89	176.25	108.25	27.87	0.46
	SFO	8.59E-11	3.22E-06	3.89E-07	7.77E-07	0.97
	SFOR	**0**	**0**	**0**	**0**	0.21
f_6	SSA	2.60E-05	3.09	1.98	0.81	0.36
	CSA	2.58	5.79	3.66	0.86	0.17
	MVO	1.46E-01	4.39	1.30	8.85E-01	0.49
	SFO	5.94E-08	1.91E-05	6.06E-06	5.37E-06	1.06
	SFOR	**8.88E-16**	**8.88E-16**	**8.88E-16**	**0**	0.26
f_7	SSA	4.34E-08	6.59E-02	1.28E-02	1.40E-02	0.37
	CSA	1.29E-01	4.79E-01	2.39E-01	8.58E-02	0.20
	MVO	3.36E-01	7.11E-01	5.60E-01	9.62E-02	0.52
	SFO	**0**	6.92E-11	7.99E-12	1.65E-11	1.07
	SFOR	**0**	**0**	**0**	**0**	0.32

SFOR algorithm were more accurate than the statistical results of the SSA, CSA, SFO, and MVO algorithms. The SFOR algorithm converged to 0 for f_5 and f_7. Although the SFOR algorithm did not converge to the optimal value 0 for f_6, the convergence values were the most competitive. For the three multimodal benchmark functions, the SFOR algorithm obtained more accurate convergence values, indicating that the SFOR algorithm effectively jumped out of the local extremum solution during the solution process, thereby converging to an accurate result. The statistical results proved that the SFOR algorithm exhibited a strong ability to avoid local extreme points. This is because the random disturbance strategy enables the SFOR algorithm to ensure the diversity of the sailfish school and the sardine school in the iterative process, which improves the ability of the SFOR algorithm to jump out of the local extreme points.

For the average running time, compared with the existing algorithms, the results of SFOR algorithm are still competitive. Whether for unimodal test function or multi-modal test function, the running time consumed by SFOR algorithm is significantly smaller than that of SFO algorithm, which indicates that SFOR algorithm can effectively solve various optimization problems and has higher efficiency. In addition, compared with SFO algorithm, SFOR algorithm does not increase the complexity of the algorithm while improving the solution effect. This is because the disturb strategy improves the diversity of the population and accelerates the convergence speed, so that SFOR algorithm can converge to the optimal solution at a faster speed.

Table 7
Results of Wilcoxon rank-sum.

f	SSA		CSA		MVO		SFO	
	P	R	P	R	P	R	P	R
f_1	3.01E-11	+	3.01E-11	+	3.01E-11	+	3.01E-11	+
f_2	3.01E-11	+	3.01E-11	+	3.01E-11	+	3.01E-11	+
f_3	3.01E-11	+	3.01E-11	+	3.01E-11	+	3.01E-11	+
f_4	3.01E-11	+	3.01E-11	+	3.01E-11	+	1.24E-04	+
f_5	1.21E-12	+	1.21E-12	+	1.21E-12	+	1.21E-12	+
f_6	1.21E-12	+	1.21E-12	+	1.21E-12	+	1.21E-12	+
f_7	1.21E-12	+	1.21E-12	+	1.21E-12	+	1.21E-12	+

The effectiveness of the SFOR algorithm was proven by a set of benchmark functions, so this study employed the SFOR-based optimization strategy to address the random disturbance problem of ELM parameters on forecasting results.

In this study, the widely used Wilcoxon rank-sum test was used, and the results obtained by each algorithm independently solving 7 test functions for 30 times are used as samples to judge the significant difference between the results of the other four comparison algorithms and the results obtained by SFOR. The test P value is set to 0.05. When the

Table 8
Pants sales forest results (billion Yuan).

Date	REAL	ELM	LSTM	SSA-ELM	SFO-ELM	SFOR-ELM
1/1/2018	2.53	2.56	2.48	2.62	2.68	2.66
1/2/2018	0.96	1.22	0.85	0.94	0.94	0.96
1/3/2018	2.52	2.53	2.59	2.42	2.55	2.57
1/3/2018	2.63	2.64	2.69	2.58	2.69	2.66
1/3/2018	3.32	3.35	3.26	3.61	3.70	3.55
1/4/2018	3.26	3.24	3.18	3.36	3.29	3.25
1/7/2018	2.72	2.73	2.72	2.51	2.61	2.58
1/8/2018	2.33	2.34	2.34	2.28	2.19	2.35
1/9/2018	2.62	2.63	2.62	2.59	2.64	2.59
1/10/2018	3.71	3.66	3.58	3.87	3.89	3.81
1/11/2018	5.26	4.71	4.49	5.06	5.14	5.25
1/12/2018	4.77	4.38	4.21	4.70	4.62	4.65
1/1/2019	3.14	3.12	3.04	3.13	2.95	2.98
1/2/2019	1.50	1.66	1.46	1.56	1.57	1.49
1/3/2019	3.05	3.02	3.03	3.03	2.88	2.97
1/4/2019	3.12	3.10	3.10	3.17	3.09	3.10

test value is<0.05, the hypothesis is rejected, indicating that the comparison algorithm has a significant difference. Otherwise, the hypothesis is accepted, indicating that the overall optimization ability of the comparison algorithm is the same. The test results of Wilcoxon rank-sum test results are shown in Table 7.

As shown in Table 7, R is "+", which means that the performance of the SFOR algorithm is better than that of the comparison algorithm. There are many cases where the P value is much<0.05, indicating that there is a statistically significant difference between the performance of the SFOR algorithm and other comparison algorithms. The results showed that the SFOR algorithm converged better than the comparison algorithms.

4.2. Case 1: Pants sales forest

On the basis of measured datasets, the feasibility and effectiveness of the SFOR-ELM model were proved through predicting the online sales of clothing products on an e-commerce platform, including pants sales data set, dress sales data set, and sweater sales data set. Appendix A lists the detailed statistical data. The statistical data from 1/5/2016 to 1/12/2017 was used as the training set, and the statistical data from 1/1/2018 to 1/4/2019 was used as the testing set. Both training data and test data include sales and features X4, X8, X9, X10. To verify the SFOR-ELM model, the predictive results of the SFOR-ELM model were compared with those of the ELM, LSTM, SSA-ELM, and SFO-ELM models.

First, the SFOR-ELM model was validated using the pants sales data set. The pants sales data set from 1/5/2016 to 1/12/2017 was used as the training set, and the pants sales data set from 1/1/2018 to 1/4/2019 was used as the test set. In the LSTM prediction process, the pants sales in the current month are predicted based on the pants sales in the first five months of the current month. From 1/1/2018 to 1/4/2019, the pants sales forecasting results of ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models are presented in Table 8 and Fig. 5.

Table 8 presented the actual values of pants sales in different periods and the predictive values of ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models. Fig. 5 showed the predictive curves of the ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models. The predictive sales value of each model fitted well with the actual value of pants sales. In addition, the SFOR-ELM, SSA-ELM, and SFO-ELM models exhibited the highest degree of approximation to the actual value curve compared with ELM model's forecasting curve. Fig. 5 indicated the sales of pants peaked in May and November. Businesses can set reasonable sales targets according to the sales trend of pants and guide the operation background to make reasonable resource allocation in advance, so as to reduce operation costs and improve competitiveness.

For the forecasting effect of pants sales, the relative error values of pants sales forecasting results are revealed in Table 9 and Fig. 6.

Table 9 presented the relative error values, and Fig. 6 revealed the relative error curve of each model. Fig. 6 presented that the fluctuation ranges of the error curves of the SSA-ELM, LSTM, SFO-ELM, and SFOR-ELM models were controlled within [-15 %, 15 %] and relatively stable. However, ELM model's error curve fluctuated greatly, with the maximum relative error exceeding 20 %. This is because volatility of random parameters reduces the predictive stability of the ELM model, resulting in poor predictive effect of the ELM model.

The evaluation index is calculated using Eqs. (24)-(26), and the calculated results are presented in Table 10 and Fig. 7.

Table 10 presented the calculation results of the evaluation

Table 9
Relative error of pants sales forest (%).

Date	ELM	LSTM	SSA-ELM	SFO-ELM	SFOR-ELM
1/1/2018	1.43	2.48	3.69	6.13	5.34
1/2/2018	26.57	0.85	−1.59	−2.22	0.23
1/3/2018	0.30	2.59	−4.15	0.91	1.91
1/3/2018	0.69	2.69	−1.86	2.28	1.22
1/3/2018	0.88	3.26	8.64	11.35	6.97
1/4/2018	−0.60	3.18	3.17	0.90	−0.31
1/7/2018	0.50	2.72	−7.69	−3.98	−5.09
1/8/2018	0.39	2.34	−2.42	−5.97	0.92
1/9/2018	0.43	2.62	−1.08	0.74	−1.02
1/10/2018	−1.22	3.58	4.32	4.95	2.81
1/11/2018	−10.45	4.49	−3.64	−2.16	−0.22
1/12/2018	−8.11	4.21	−1.37	−3.06	−2.40
1/1/2019	−0.54	3.04	−0.41	−6.05	−5.18
1/2/2019	10.53	1.46	3.86	4.40	−1.17
1/3/2019	−1.12	3.03	−0.76	−5.48	−2.63
1/4/2019	−0.63	3.10	1.75	−0.89	−0.40

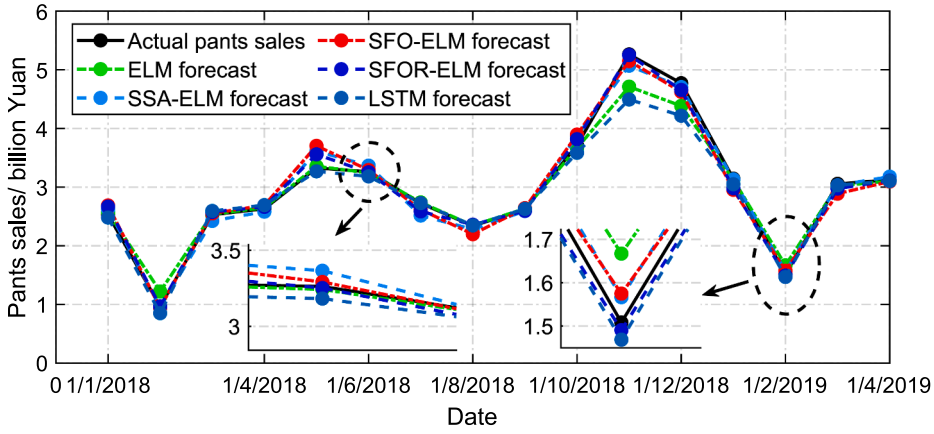

Fig. 5. Pants sales forecast results.

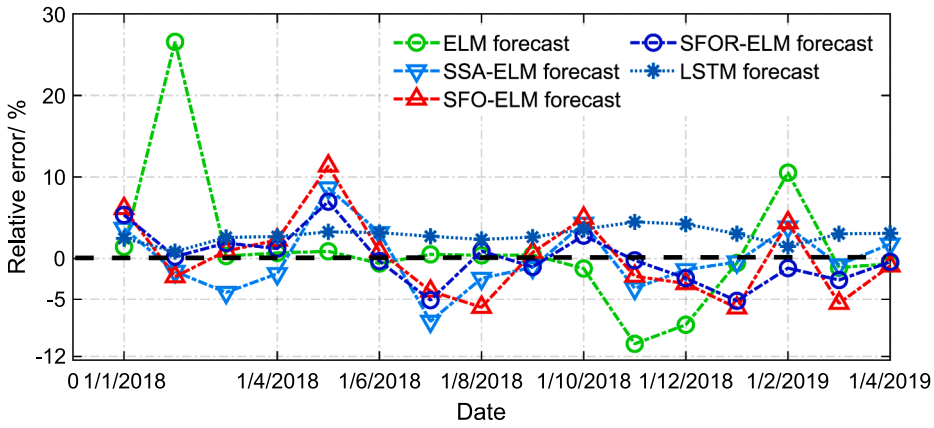

Fig. 6. Relative error curves of pants sales forest.

Table 10
Evaluation results for pants sales forecast.

Approach	Evaluation indicators/ %		
	MAPE	*RMSE*	*R2*
ELM	4.02	18.55	99.43
LSTM	3.76	24.46	97.26
SSA-ELM	3.15	12.19	98.56
SFO-ELM	3.84	14.61	97.94
SFOR-ELM	2.36	9.81	99.07

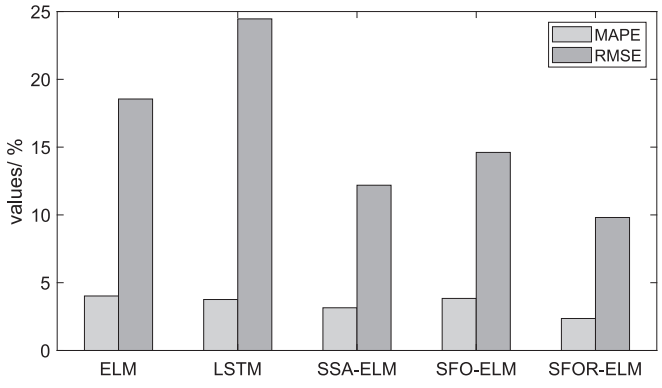

Fig. 7. Comparison of evaluation results.

indicators of the ELM, SSA-ELM, SFO-ELM, and SFOR-ELM models. Fig. 7 presented the comparison between RMSE and MAPE of each model. The evaluation result of the ELM model was the worst, which could also be seen in the forecasting curves in Figs. 5 and 6. The RMSE value of the ELM model exceeded 15 %, but the R2 value was the highest, reaching 99.43 %, indicating that ELM model's forecast curve fitted the actual value curve to a high degree, but the prediction error was relatively large. The SFOR-ELM model achieved satisfactory predictive results. Compare with the ELM, LSTM, SFO-ELM, and SFO-ELM models, the MAPE of the SFOR-ELM model was reduced by 1.66 %, 1.4 %, 0.79 %, and 1.48 %. The RMSE value was reduced by 8.74 %, 14.65 %, 2.28 %, and 4.8 %, respectively.

4.3. Case 2: Dress sales forest

Second, the SFOR-ELM model was validated using the dress sales data set. The dress sales data set from 1/5/2016 to 1/12/2017 was used as the training set, and the dress sales data set from 1/1/2018 to 1/4/2019 was used as the test set. In the LSTM prediction process, the dress sales in the current month are predicted based on the dress sales in the first five months of the current month. From 1/1/2018 to 1/4/2019, the dress sales forecast results of ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models are shown in Table 11 and Fig. 8.

Table 11
Dress sales forest results (billion Yuan).

Date	Real	ELM	LSTM	SSA-ELM	SFO-ELM	SFOR-ELM
1/1/2018	2.20	2.29	2.25	2.21	2.23	2.18
1/2/2018	1.14	1.40	1.26	1.14	1.14	1.14
1/3/2018	3.87	4.27	4.32	3.73	3.85	3.72
1/3/2018	4.71	5.07	5.24	4.93	5.00	4.86
1/3/2018	6.96	6.99	7.37	7.15	7.19	7.16
1/4/2018	7.82	7.47	7.79	7.36	7.31	7.38
1/7/2018	6.37	6.34	6.57	6.75	6.66	6.66
1/8/2018	4.57	4.68	4.71	4.56	4.30	4.43
1/9/2018	3.64	3.76	3.69	3.64	3.63	3.70
1/10/2018	3.18	3.35	3.34	3.37	3.44	3.24
1/11/2018	3.77	3.95	3.87	3.79	3.79	3.77
1/12/2018	3.30	3.50	3.45	3.31	3.29	3.28
1/1/2019	3.28	3.45	3.41	3.18	3.15	3.30
1/2/2019	2.37	2.73	2.70	2.34	2.31	2.37
1/3/2019	5.21	5.52	5.66	5.01	5.07	5.14
1/4/2019	6.28	6.41	6.71	6.22	6.32	6.33

Table 11 presented the actual values of dress sales in different periods and the predicted values of ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models. Fig. 8 showed the prediction curves of the ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models. In most periods, the fit degree between the predicted value of the dress sales of each model and the actual value of the dress sales was high. However, in the period of 1/6/2018, the fitting effect was poor. Additionally, similar to the forecasting results in Fig. 5, although the dress sales forecast curve of the ELM model fitted the changing trend of the actual dress sales, there was a large deviation between the forecasting value and the actual value. Fig. 8 showed the sales of dresses peaked in June. Fig. 8 showed the sales of dresses peaked in June. Merchants can formulate corresponding sales strategies according to the sales trend of dresses. The operating costs are reduced and the competitiveness of goods is improved through rational allocation of sales resources. The relative error values of dress forecasting results in Table 12 and Fig. 9.

Table 12 presented the relative error values of dress sales foresting results. Fig. 9 showed the relative error curve of each model. The error curve fluctuation ranges of the SSA-ELM, SFO-ELM, and SFOR-ELM models were stable and controlled within [-10 %, 10 %]. Similar to the prediction results in Fig. 6, ELM model's error curve fluctuated greatly due to the influence of random parameters, with the relative errors exceeding 5 % in most periods and the maximum relative error exceeding 20 %. In contrast, the prediction error fluctuations of SSA-ELM, SFO-ELM, and SFOR-ELM models were smaller by optimizing the parameters. In addition, the prediction errors of LSTM model are

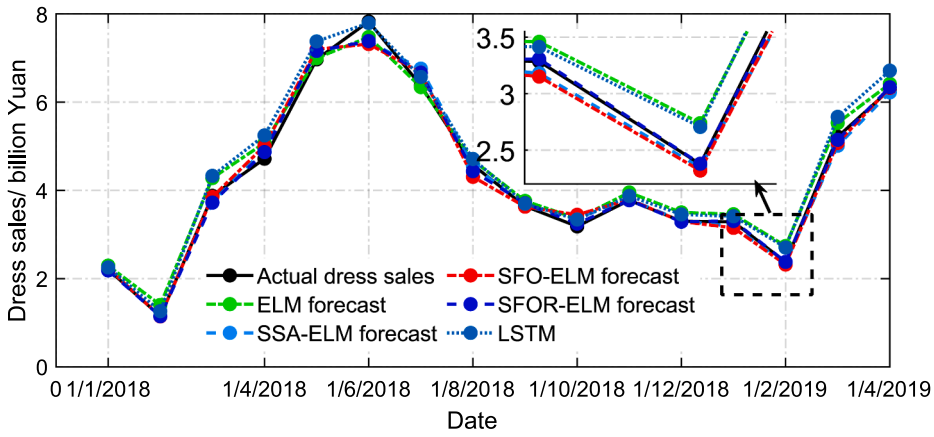

Fig. 8. Dress sales forecast results.

Table 12
Relative error of dress sales forest (%).

Date	ELM	LSTM	SSA-ELM	SFO-ELM	SFOR-ELM
1/1/2018	3.77	2.01	0.24	1.15	−1.07
1/2/2018	22.42	10.17	0.02	−0.20	0.18
1/3/2018	10.42	11.72	−3.74	−0.62	−3.95
1/3/2018	7.53	11.20	4.61	6.18	3.18
1/3/2018	0.42	5.84	2.73	3.26	2.77
1/4/2018	−4.50	−0.43	−5.86	−6.52	−5.66
1/7/2018	−0.51	3.06	6.01	4.59	4.56
1/8/2018	2.39	2.99	−0.28	−5.85	−3.02
1/9/2018	3.26	1.55	0.04	−0.30	1.76
1/10/2018	5.46	4.91	5.86	8.26	1.94
1/11/2018	4.62	2.58	0.55	0.42	−0.01
1/12/2018	6.02	4.64	0.31	−0.23	−0.35
1/1/2019	5.26	3.89	−3.20	−4.10	0.64
1/2/2019	15.29	13.89	−1.23	−2.37	0.16
1/3/2019	6.01	8.63	−3.89	−2.68	−1.32
1/4/2019	2.04	6.79	−0.99	0.60	0.82

large, which indicates that the prediction stability of the model is poor.

The calculated results of evaluation indexes are shown in Table 13 and Fig. 10.

Table 13 presented the evaluation results of ELM, LSTM, SSA-ELM, SFO-ELM, and SFOR-ELM models for predicting dress sales. Fig. 10 presented the comparison between RMSE and MAPE of each model. The evaluation results revealed that the ELM model did not get satisfactory evaluation results due to the influence of random parameters. The predictive curves in Figs. 8 and 9 also showed a large deviation between the predicted and the actual values. However, the fitting effect of the ELM model was better, and its R2 value was the most competitive, reaching 99.31 %. The evaluation results of the SFOR-ELM model were the most competitive. Compared with the ELM, LSTM, SFO-ELM, and SFO-ELM models, the MAPE of the SFOR-ELM model was reduced by 4.29 %, 3.93 %, 1 %, and 0.51 %, in addition, the RMSE value was reduced by 7.8 %, 12.65 %, 4.62 %, 2.81 %. Meanwhile, the fitting effect of the SFOR-ELM model was competitive, and its R2 reached 99.22 %.

4.4. Case 3: Sweater sales forest

Finally, the SFOR-ELM model was validated using the sweater sales data set. The sweater sales data set from 1/5/2016 to 1/12/2017 was used as the training set, and the sweater sales data set from 1/1/2018 to 1/4/2019 was used as the test set. In the LSTM prediction process, the sweater sales in the current month are predicted based on the sweater sales in the first five months of the current month. The sweater sales forecast results of ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models from 1/1/2018 to 1/4/2019 were shown in Table 14 and Fig. 11.

Table 13
Evaluation results for dress sales forecast.

Approach	Evaluation indicators/ %		
	MAPE	*RMSE*	*R2*
ELM	6.25	23.62	99.31
LSTM	5.89	28.47	99.18
SSA-ELM	2.96	20.44	98.69
SFO-ELM	2.47	18.63	98.91
SFOR-ELM	1.96	15.82	99.22

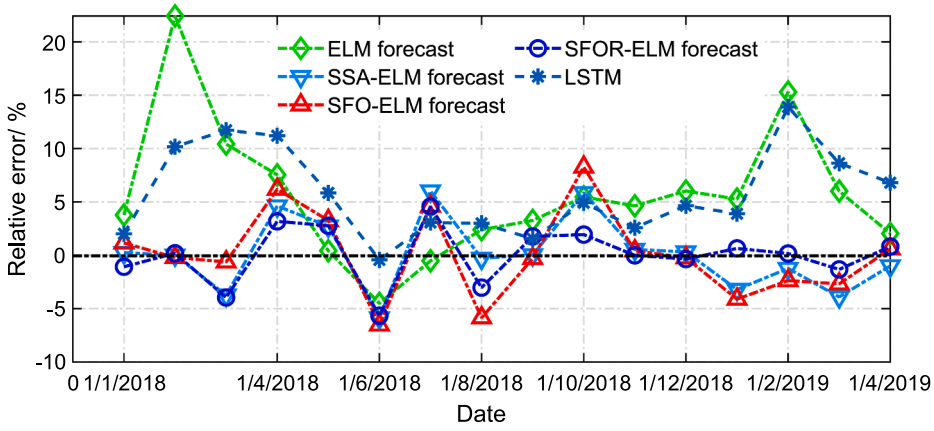

Fig. 9. Relative error curves of dress sales forest.

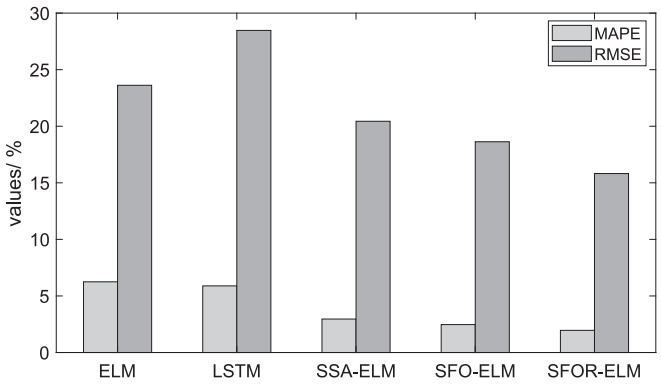

Fig. 10. Comparison of evaluation results for dress sales.

Table 14
Sweater sales forest results (billion Yuan).

Date	Real	ELM	LSTM	SSA-ELM	SFO-ELM	SFOR-ELM
1/1/2018	1.43	1.21	1.53	1.32	1.41	1.38
1/2/2018	0.63	0.57	1.01	0.63	0.62	0.68
1/3/2018	1.32	1.11	1.46	1.30	1.51	1.35
1/3/2018	0.80	0.69	0.93	0.79	0.77	0.82
1/3/2018	0.67	0.63	0.76	0.69	0.70	0.68
1/4/2018	0.64	0.60	0.80	0.64	0.64	0.62
1/7/2018	0.63	0.58	0.87	0.60	0.61	0.64
1/8/2018	1.29	0.99	1.66	1.20	1.08	1.41
1/9/2018	2.34	2.30	2.57	2.83	2.76	2.41
1/10/2018	3.28	2.87	3.69	3.32	3.39	3.09
1/11/2018	3.89	2.95	3.80	3.36	3.39	3.37
1/12/2018	2.93	2.73	2.86	3.02	2.92	2.71
1/1/2019	2.16	1.98	2.15	2.19	2.21	1.98
1/2/2019	1.00	0.77	1.38	1.10	1.16	1.04
1/3/2019	1.90	1.70	1.97	1.97	1.92	1.82
1/4/2019	1.15	0.95	1.27	1.07	0.94	1.14

Table 14 presented the actual values of sweater sales in different periods and the predicted values of ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models. Fig. 11 presented the predictive results of sweater sales. For most of the time periods, the predicted value and actual value of each model exhibited a better fitting effect, however, in the 1/11/2018 time period, the fitting effect was poor. From an overall point of view, the prediction curves of the ELM, LSTM, SFOR-ELM, SSA-ELM, and SFO-ELM models well reflected the fluctuation trend of the actual sweater sales curve. In addition, the predictive curve of the SFOR-ELM model fitted well with the true value curve of sweater sales. Fig. 11 indicated the sales of sweaters peaked in November. Merchants can set corresponding sales targets according to the sales trend of sweaters. By mining a large amount of information contained in sweater sales historical data, we can make favorable decisions to deal with the fierce competition.

The relative error values of online clothing sales forecasting methods were presented in Table 15 and Fig. 12.

Table 15 presented the relative error values of the four online clothing sales forecasting approaches. Fig. 12 revealed the relative error curve of each model. The relative error of the SSA-ELM and SFO-ELM models was basically controlled in the interval [-20 %, 20 %]. The error curve of the SFOR-ELM model exhibited the smallest fluctuation range, the relative error value was basically controlled within [-10 %, 10 %], and the curve fluctuation range was relatively stable. The relative error of the ELM model fluctuated greatly, and most of the relative error values were in the interval [-10 %, 30 %]. For the LSTM model, the maximum relative error exceeds 60 %, indicating that the model can not accurately reflect the fluctuation trend of sweater sales series.

According to equations (24), (25) and (26), the sweater sales forecast results were evaluated. The calculated evaluation results of each model are shown in Table 16 and Fig. 13.

Table 16 showed the evaluation results of the ELM, LSTM, SSA-ELM, SFO-ELM, and SFOR-ELM models for predicting sweater sales. Fig. 13 presented the comparison results of RMSE and MAPE for each model. The RMSE and MAPE of the ELM model were the highest, reaching 30.21 % and 12.37 % respectively. Figs. 11 and 12 showed a large error between the forecast curve of sweater sales and the true value curve. The SFOR-ELM model obtained the smallest MAPE and RMSE values, which were 5.03 % and 16.19 %, respectively. Compared with the ELM, LSTM, SFO-ELM, and SFO-ELM models, the MAPE of the SFOR-ELM model was reduced by 7.34 %, 12.59 %, 0.43 %, and 2.21 %. Additionally, the

Table 15
Relative error of sweater sales forest (%).

Date	ELM	LSTM	SSA-ELM	SFO-ELM	SFOR-ELM
1/1/2018	−15.14	6.77	−7.62	−1.13	−3.56
1/2/2018	−8.59	60.80	0.03	−1.14	8.89
1/3/2018	−15.23	11.02	−1.33	14.92	2.36
1/3/2018	−12.93	16.86	−1.38	−3.00	2.82
1/3/2018	−6.47	13.55	2.96	4.38	1.48
1/4/2018	−5.44	25.31	1.02	0.25	−3.28
1/7/2018	−6.71	38.98	−3.45	−2.40	2.15
1/8/2018	−23.30	28.92	−7.21	−16.50	9.65
1/9/2018	−1.61	9.70	20.90	17.69	3.08
1/10/2018	−12.37	12.58	1.28	3.49	−5.66
1/11/2018	−24.21	−2.45	−13.70	−13.00	−13.32
1/12/2018	−6.71	−2.31	3.37	−0.15	−7.45
1/1/2019	−8.12	−0.26	1.74	2.69	−8.16
1/2/2019	−23.07	37.89	10.32	15.52	3.63
1/3/2019	−10.54	3.95	3.77	1.36	−3.93
1/4/2019	−17.49	10.52	−7.27	−18.21	−1.06

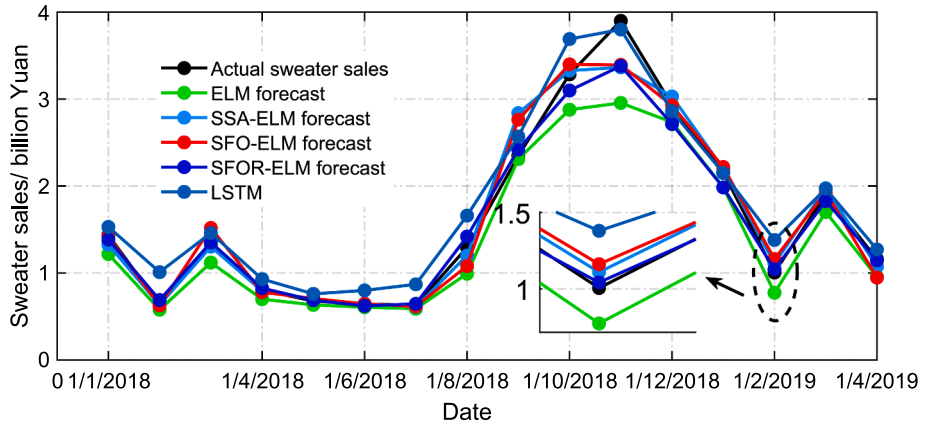

Fig. 11. Sweater sales forecast results.

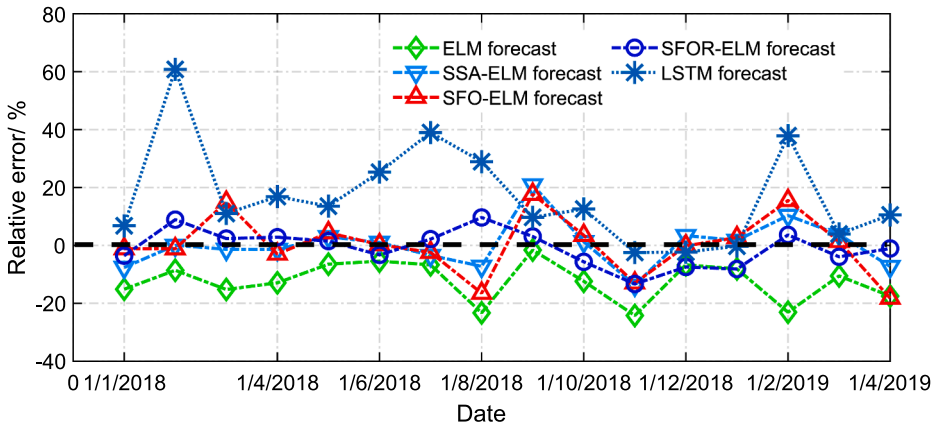

Fig. 12. Relative error curves of sweater sales forest.

Table 16
Evaluation results for sweater sales forecast.

Approach	Evaluation indicators/ %		
	MAPE	*RMSE*	R^2
ELM	12.37	30.21	97.09
LSTM	17.62	22.84	97.71
SSA-ELM	5.46	19.11	96.32
SFO-ELM	7.24	19.40	96.20
SFOR-ELM	5.03	16.19	98.98

RMSE value of the SFOR-ELM model was reduced by 14.02 %, 6.65 %, 2.92 %, and 3.21 %, respectively. At the same time, the R2 of the SFOR-ELM model was the highest, reaching 98.98 %, indicating that the prediction curve of the SFOR-ELM model accurately reflected the changing trend of the actual curve of sweater sales.

In this section, three different cases were applied to verify the SFOR-ELM model. The test results proved that the SFOR-ELM model proposed in this research was suitable for multi-scenario and multi-product sales forecasts. The accurate forecast of online clothing sales plays an important role in updating the sales strategy of e-commerce platform and improving the management efficiency and economic benefits of e-commerce platform.

5. Concluding remarks

At present, information technology has penetrated into all fields of economic life, and has promoted the transformation of the traditional economic development mode based on material production and material services to the new economic development mode based on information production and information services. The rapid development of the e-commerce economy based on information technology has become the driving force promoting economic growth. Commodity sales forecasts affect the sales plan and inventory stability of e-commerce platforms. Therefore, accurate commodity sales forecasting is critical to improve the management efficiency and economic benefits of e-commerce platforms. To this end, this research proposed a novel online sales forecast approach for clothing products based on a data mining technique, and proved the proposed method using the statistical data of an e-commerce platform. The contributions and conclusions of this research are as follows:

- To improve the convergence performance of the SFO algorithm, an SFO algorithm based on a random disturbance strategy is proposed, and the convergence effect of the SFOR algorithm is analysed using a set of benchmark functions. The test results reveal that the SFOR algorithm exhibits the highest convergence accuracy compared to the state-of-the-art heuristic optimization algorithms for the unimodal benchmark function. In addition, the SFOR algorithm

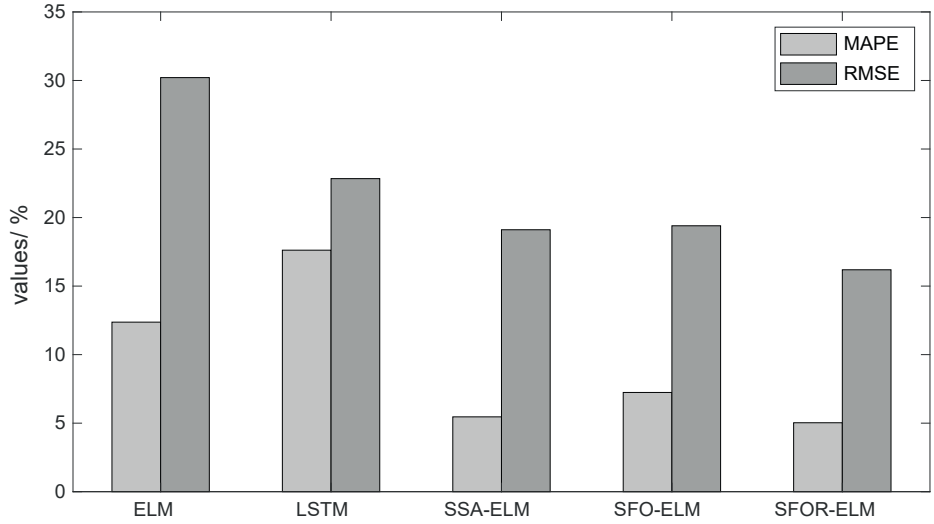

Fig. 13. Comparison of evaluation results for sweater sales.

Table A1
Statistical data for pants sales.

Date	S/ billion Yuan	X4/ billion	X8/ 10 million	X9/ 100 thousand	X10/ million
1/5/ 2016	2.23	3.84	1.48	1.12	1.19
1/6/ 2016	2.14	3.85	1.44	1.09	1.16
1/7/ 2016	1.72	3.30	1.19	0.97	1.03
1/8/ 2016	1.46	2.81	1.00	0.87	0.93
1/9/ 2016	1.86	3.05	1.17	1.02	1.07
1/10/ 2016	2.47	4.20	1.71	1.25	1.26
1/11/ 2016	3.80	5.62	2.44	1.49	1.63
1/12/ 2016	3.03	5.75	1.80	1.29	1.42
1/1/ 2017	1.77	3.59	0.98	0.92	1.04
1/2/ 2017	1.46	3.34	1.00	0.82	0.93
1/3/ 2017	2.26	4.78	1.45	1.05	1.20
1/4/ 2017	2.42	5.53	1.68	1.14	1.25
1/5/ 2017	2.81	6.84	2.02	1.24	1.36
1/6/ 2017	2.78	6.62	1.92	1.18	1.35
1/7/ 2017	2.26	5.71	1.64	1.08	1.20
1/8/ 2017	1.98	4.73	1.42	0.98	1.11
1/9/ 2017	2.34	4.62	1.49	1.07	1.23
1/10/ 2017	3.31	6.44	2.27	1.34	1.50
1/11/ 2017	3.50	5.70	2.12	1.19	1.55
1/12/ 2017	2.95	4.80	1.65	1.07	1.40
1/1/ 2018	2.53	4.17	1.33	0.98	1.28
1/2/ 2018	0.96	1.99	0.69	0.58	0.74
1/3/ 2018	2.52	4.49	1.68	0.97	1.28
1/4/ 2018	2.63	5.12	1.84	1.00	1.31
1/5/ 2018	3.32	6.66	2.35	1.13	1.50
1/6/ 2018	3.26	6.31	2.10	1.09	1.49
1/7/ 2018	2.72	5.72	1.83	1.01	1.34
1/8/ 2018	2.33	4.50	1.48	0.90	1.22
1/9/ 2018	2.62	4.25	1.52	0.99	1.31
1/10/ 2018	3.71	6.16	2.29	1.17	1.60
1/11/ 2018	5.26	7.63	3.00	1.37	1.97
1/12/ 2018	4.77	6.97	2.58	1.29	1.86
1/1/ 2019	3.14	4.86	1.61	1.05	1.45
1/2/ 2019	1.50	2.95	1.04	0.69	0.95
1/3/ 2019	3.05	5.06	1.81	1.01	1.43
1/4/ 2019	3.12	5.45	1.95	1.04	1.45

Table A2
Statistical data for dress sales.

Date	S/ billion Yuan	X4/ billion	X8/ 10 million	X9/ 100 thousand	X10/ million
1/5/ 2016	5.29	11.38	3.03	1.21	1.98
1/6/ 2016	5.66	12.45	3.21	1.24	2.06
1/7/ 2016	4.87	11.70	2.89	1.18	1.88
1/8/ 2016	3.66	8.41	2.06	0.97	1.59
1/9/ 2016	3.35	7.24	1.87	0.92	1.51
1/10/ 2016	2.67	6.42	1.70	0.84	1.32
1/11/ 2016	2.82	5.61	1.65	0.81	1.36
1/12/ 2016	2.56	6.85	1.59	0.77	1.29
1/1/ 2017	1.98	6.12	1.28	0.63	1.11
1/2/ 2017	2.35	7.49	1.68	0.72	1.23
1/3/ 2017	3.70	10.24	2.32	0.94	1.60
1/4/ 2017	4.33	13.09	2.90	1.06	1.76
1/5/ 2017	6.42	19.23	4.07	1.31	2.22
1/6/ 2017	6.85	19.10	4.01	1.31	2.31
1/7/ 2017	5.57	16.64	3.44	1.20	2.04
1/8/ 2017	3.86	10.83	2.31	0.94	1.64
1/9/ 2017	3.41	8.45	1.93	0.86	1.52
1/10/ 2017	2.62	6.78	1.64	0.75	1.31
1/11/ 2017	2.33	4.16	1.07	0.58	1.22
1/12/ 2017	1.81	3.94	0.95	0.50	1.06
1/1/ 2018	2.20	5.07	1.20	0.56	1.18
1/2/ 2018	1.14	3.60	0.94	0.38	0.81
1/3/ 2018	3.87	8.80	2.39	0.77	1.64
1/4/ 2018	4.71	11.27	2.93	0.88	1.85
1/5/ 2018	6.96	16.61	4.19	1.08	2.33
1/6/ 2018	7.82	17.18	4.10	1.11	2.50
1/7/ 2018	6.37	15.59	3.55	1.04	2.21
1/8/ 2018	4.57	10.37	2.40	0.85	1.81
1/9/ 2018	3.64	7.22	1.78	0.75	1.58
1/10/ 2018	3.18	7.21	1.82	0.73	1.46
1/11/ 2018	3.77	6.71	1.81	0.76	1.62
1/12/ 2018	3.30	6.40	1.71	0.70	1.50
1/1/ 2019	3.28	6.83	1.70	0.69	1.49
1/2/ 2019	2.37	5.91	1.64	0.57	1.23
1/3/ 2019	5.21	10.70	2.90	0.88	1.96
1/4/ 2019	6.28	13.45	3.54	1.00	2.19

Table A3
Statistical data for sweater sales.

Date	S/ billion Yuan	X4/ billion	X8/ 10 million	X9/ 100 thousand	X10/ million
1/5/ 2016	0.63	1.11	3.66	4.71	5.87
1/6/ 2016	0.46	0.82	2.72	3.91	4.97
1/7/ 2016	0.42	0.81	2.70	3.55	4.69
1/8/ 2016	0.91	1.86	6.60	5.33	7.20
1/9/ 2016	1.90	3.78	12.69	8.13	10.92
1/10/ 2016	2.31	4.68	15.24	9.07	12.19
1/11/ 2016	2.74	4.26	15.65	9.14	13.48
1/12/ 2016	2.02	4.22	12.13	7.90	11.30
1/1/ 2017	1.27	3.04	7.69	6.09	8.68
1/2/ 2017	1.06	2.75	7.08	5.58	7.82
1/3/ 2017	1.45	3.49	8.85	6.77	9.36
1/4/ 2017	0.89	2.31	5.83	5.48	7.08
1/5/ 2017	0.62	1.50	4.01	4.54	5.80
1/6/ 2017	0.55	1.18	3.32	4.12	5.43
1/7/ 2017	0.51	1.31	3.89	3.96	5.25
1/8/ 2017	1.27	2.91	9.74	6.11	8.68
1/9/ 2017	2.21	5.02	15.47	8.40	11.88
1/10/ 2017	2.98	6.55	20.26	9.93	14.14
1/11/ 2017	2.42	3.71	12.46	7.63	12.54
1/12/ 2017	1.57	2.72	8.49	6.03	9.76
1/1/ 2018	1.43	2.69	7.86	5.78	9.27
1/2/ 2018	0.63	1.55	4.62	3.62	5.86
1/3/ 2018	1.32	2.61	8.05	5.73	8.85
1/4/ 2018	0.80	1.65	5.21	4.64	6.68
1/5/ 2018	0.67	1.30	4.30	4.39	6.08
1/6/ 2018	0.64	1.09	3.56	4.03	5.90
1/7/ 2018	0.63	1.19	3.91	3.87	5.85
1/8/ 2018	1.29	2.29	8.32	5.24	8.74
1/9/ 2018	2.34	4.08	13.94	7.48	12.30
1/10/ 2018	3.28	5.97	19.61	8.93	14.96
1/11/ 2018	3.89	5.51	19.26	9.33	16.55
1/12/ 2018	2.93	4.17	14.26	7.85	14.00
1/1/ 2019	2.16	3.42	10.31	6.67	11.72
1/2/ 2019	1.00	1.97	6.24	4.45	7.58
1/3/ 2019	1.90	3.16	9.85	6.44	10.90
1/4/ 2019	1.15	2.24	6.75	5.34	8.21

effectively avoids local extremum solutions for the multimodal benchmark functions.

- The constructed SFOR-ELM model was employed to forecast the online sales of clothing products. For three different cases, the SFOR-ELM model achieves satisfactory predictive results. For Case 1, the MAPE of the SFOR-ELM model is 1.66 %, 1.4 %, 0.79 %, and 1.48 % smaller than those of the ELM, SFO-ELM, and SFO-ELM models, respectively. Meanwhile, the RMSE value of the SFOR-ELM model is 8.74 %, 14.65 %, 2.28 %, and 4.8 % smaller than the RMSE of the ELM, SFO-ELM, and SFO-ELM models, respectively.
- The presented online sales forecast approach of clothing products based on data mining technique provides a new idea for the accurate forecasting of multiscene and multicommodity sales, and provides a theoretical basis for improving the management efficiency and economic benefits of e-commerce platforms.

Although the proposed online sales forecast approach achieves competitive results, this study still has the following limitations. First, the SFOR algorithm needs to further develop the solving ability. Second, the generalization ability and scalability of the online sales forecast approach need to be further enhanced. We will focus on solving the abovementioned limitations in future research.

CRediT authorship contribution statement

Bo Zhang: Conceptualization. **Ming-Lang Tseng:** Conceptualization. **Lili Qi:** Conceptualization. **Yuehong Guo:** Conceptualization. **Ching-Hsin Wang:** Conceptualization.

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.

Data availability

The data that has been used is confidential.

Acknowledgement

This study is partially supported by ministry of Science and Technology, Taiwan MOST- 110-2221-E-468 -010.

Appendix A

Tables A1-A3.

References

Albadr, M. A. A., & Tiun, S. (2020). Spoken Language Identification Based on Particle Swarm Optimisation-Extreme Learning Machine Approach. *Circuits Systems and Signal Processing, 39*(9), 4596–4622.

Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. *Computers & Structures, 169*, 1–12.

Bag, S., Gupta, S., & Kumar, S. (2021). Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development. *International Journal of Production Economics, 231*.

Boysen, N., de Koster, R., & Weidinger, F. (2019). Warehousing in the e-commerce era: A survey. *European Journal of Operational Research, 277*(2), 396–411.

Chai, S. L., Zhang, Z. X., & Zhang, Z. (2021). Carbon price prediction for China's ETS pilots using variational mode decomposition and optimized extreme learning machine. *Annals of Operations Research, 22*.

Chen, Z., Zhao, H. Q., Zhang, Y. J., Shen, S. Q., Shen, J. W., & Liu, Y. G. (2022). State of health estimation for lithium-ion batteries based on temperature prediction and gated recurrent unit neural network. *Journal of Power Sources, 521*, 13.

Choi, J. Y., & Lee, B. (2018). Combining LSTM Network Ensemble via Adaptive Weighting for Improved Time Series Forecasting. *Mathematical Problems in Engineering, 2018*, 8.

Correia, A., Lopes, C., Silva, E. C. E., Monteiro, M., & Lopes, R. B. (2020). A multi-model methodology for forecasting sales and returns of liquefied petroleum gas cylinders. *Neural Computing & Applications, 32*(16), 12643–12669.
Das, S., Sahu, T. P., Janghel, R. R., & Sahu, B. K. (2022). Effective forecasting of stock market price by using extreme learning machine optimized by PSO-based group oriented crow search algorithm. *Neural Computing & Applications, 34*(1), 555–591.
El Hammouti, I., Lajjam, A., El Merouani, M., & Tabaa, Y. (2019). A modified sailfish optimizer to solve dynamic berth allocation problem in conventional container terminal. *International Journal of Industrial Engineering Computations, 10*(4), 491–504.
El Ouadi, J., Malhene, N., Benhadou, S., & Medromi, H. (2022). Towards a machine-learning based approach for splitting cities in freight logistics context: Benchmarks of clustering and prediction models. *Computers & Industrial Engineering, 166*.
Giri, C., Thomassey, S., & Zeng, X. Y. (2019). Exploitation of Social Network Data for Forecasting Garment Sales. *International Journal of Computational Intelligence Systems, 12*(2), 1423–1435.
Gautam, A. S. S., & Singh, S. R. (2020). A new method of time series forecasting using intuitionistic fuzzy set based on average-length. *Journal of Industrial and Production Engineering, 37*(4), 175–185.
Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501.
Islam, M. J., Xia, Y. Y., & Sattar, J. (2020). Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robotics and Automation Letters, 5*(2), 3227–3234.
Ji, S. W., Wang, X. J., Zhao, W. P., & Guo, D. (2019). An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise. *Mathematical Problems in Engineering, 2019*, 15.
Karasu, S., Altan, A., Bekiros, S., & Ahmad, W. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy, 212*, 12.
Kardani, N., Bardhan, A., Gupta, S., Samui, P., Nazem, M., Zhang, Y., & Zhou, A. (2021). Predicting permeability of tight carbonates using a hybrid machine learning approach of modified equilibrium optimizer and extreme learning machine. *Acta Geotechnica*.
Lai, X. J., Zhang, S., Mao, N., Liu, J. J., & Chen, Q. X. (2022). Kansei engineering for new energy vehicle exterior design: An internet big data mining approach. *Computers & Industrial Engineering, 165*.
Larrea, M., Porto, A., Irigoyen, E., Barragan, A. J., & Andujar, J. M. (2021). Extreme learning machine ensemble model for time series forecasting boosted by PSO: Application to an electric consumption problem. *Neurocomputing, 452*, 465–472.
Li, L. L., Liu, Z. F., Tseng, M. L., Zheng, S. J., & Lim, M. K. (2021). Improved tunicate swarm algorithm: Solving the dynamic economic emission dispatch problems. *Applied Soft Computing, 108*.
Liu, Z. F., Li, L. L., Liu, Y. W., Liu, J. Q., Li, H. Y., & Shen, Q. (2021). Dynamic economic emission dispatch considering renewable energy generation: A novel multi-objective optimization approach. *Energy, 235*.
Marathe, P. S., Westerhof, R. J. M., & Kersten, S. R. A. (2019). Fast pyrolysis of lignins with different molecular weight: Experiments and modelling. *Applied Energy, 236*, 1125–1137.
Mohanty, D. K., Parida, A. K., & Suman, S. (2021). Financial market prediction under deep learning framework using auto encoder and kernel extreme learning machine. *Applied Soft Computing, 99*, 14.
Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., & Mirjalili, S. M. (2017). Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software, 114*, 163–191.
Mirjalili, S., Mirjalili, S. M., & Hatamlou, A. (2016). Multi-Verse Optimizer: A nature-inspired algorithm for global optimization. *Neural Computing & Applications, 27*(2), 495–513.
Nassef, M. G. A., Hussein, T. M., & Mokhiamar, O. (2021). An adaptive variational mode decomposition based on sailfish optimization algorithm and Gini index for fault identification in rolling bearings. *Measurement, 173*.
Okpoti, E. S., Jeong, I. J., & Moon, S. K. (2019). Decentralized determination of design variables among cooperative designers for product platform design in a product family. *Computers & Industrial Engineering, 135*, 601–614.
Sano, H., & Yamada, K. (2021). Prediction accuracy of sales surprise for inventory turnover. *International Journal of Production Research, 59*(17), 5337–5351.
Silva, A. W. B., Freitas, B. B., Filho, C. L. A., Freitas, C. D., Junior, E. A. S., Castro, E. S., … Carneiro, T. C. (2022). Methodology Based on Artificial Neural Networks for Hourly Forecasting of PV Plants Generation. *Ieee Latin America Transactions, 20*(4), 659–668.
Silitonga, A. S., Shamsuddin, A. H., Mahlia, T. M. I., Milano, J., Kusumo, F., Siswantoro, J., Dharma, S., Sebayang, A. H., Masjuki, H. H., & Ong, H. C. (2020). Biodiesel synthesis from Ceiba pentandra oil by microwave irradiation-assisted transesterification: ELM modeling and optimization. *Renewable Energy, 146*, 1278–1291.
Shao, B. L., Li, M. L., Zhao, Y., & Bian, G. Q. (2019). Nickel price forecast based on the LSTM neural network optimized by the improved PSO algorithm. *Mathematical Problems in Engineering, 2019*, 15.
Shariati, M., Mafipour, M. S., Ghahremani, B., Azarhomayun, F., Ahmadi, M., Trung, N. T., & Shariati, A. (2020). A novel hybrid extreme learning machine-grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement. *Engineering with Computers, 23*.
Shadravan, S., Naji, H. R., & Bardsiri, V. K. (2019). The Sailfish Optimizer: A novel nature-inspired metaheuristic algorithm for solving constrained engineering optimization problems. *Engineering Applications of Artificial Intelligence, 80*, 20–34.
Sun, S. L., Jin, F., Li, H. T., & Li, Y. W. (2021). A new hybrid optimization ensemble learning approach for carbon price forecasting. *Applied Mathematical Modelling, 97*, 182–205.
Tutuncu, K., Sahman, M. A., & Tusat, E. (2021). A hybrid binary grey wolf optimizer for selection and reduction of reference points with extreme learning machine approach on local GNSS/leveling geoid determination. *Applied Soft Computing, 108*.
Tseng, M. L., Tran, T. P. T., Ha, H. M., Bui, T. D., & Lim, M. K. (2021). Sustainable industrial and operation engineering trends and challenges Toward Industry 4.0: A data driven analysis, Journal of Industrial and Production. *Engineering, 38*(8), 581–598. https://doi.org/10.1080/21681015.2021.1950227
Tsoumakas, G. (2019). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review, 52*(1), 441–447.
Wankhede, J., Kumar, M., & Sambandam, P. (2020). Efficient heart disease prediction-based on optimal feature selection using DFCSS and classification by improved Elman-SFO. *IET systems biology, 14*(6), 380–390.
Wang, X., Gao, S., Zhou, S. Y., Guo, Y. B., Duan, Y. H., & Wu, D. Q. (2021). Prediction of House Price Index Based on Bagging Integrated WOA-SVR Model. *Mathematical Problems in Engineering, 2021*, 15.
Weng, T. Y., Liu, W. Y., & Xiao, J. (2020). Supply chain sales forecasting based on lightGBM and LSTM combination model. *Industrial Management & Data Systems, 120*(2), 265–279.
Wu, D. M., Wang, X. L., & Wu, S. C. (2021). A Hybrid Method Based on Extreme Learning Machine and Wavelet Transform Denoising for Stock Prediction. *Entropy, 23*(4), 30.
Xu, T. T., Peng, Z., & Wu, L. F. (2021). A novel data-driven method for predicting the circulating capacity of lithium-ion battery under random variable current. *Energy, 218*, 13.
Yahia, S., Said, S., & Zaied, M. (2020). A novel classification approach based on Extreme Learning Machine and Wavelet Neural Networks. *Multimedia Tools and Applications, 79*(19–20), 13869–13890.
Zhang, T. T., Tang, Z. P., Wu, J. C., Du, X. X., & Chen, K. J. (2021). Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. *Energy, 229*, 13.