UNDER PEER REVIEW

Original Research Article

Combined model based on LASSO and WOA for natural gas consumption forecast

.

Abstract

As the impact of the Russia-Ukraine war continues to expand, energy shortages appear in Europe. After Russia cut off the Nord Stream 1 pipeline that transported natural gas to Europe, most European countries experienced a natural gas crisis, and Germany was severely affected. In order to effectively predict the consumption of natural gas, this paper combines Least Absolute Shrinkage and Select Operator model with *Whale Optimization Algorithm*, uses the NAR model to reconstruct the phase space of the original time series, and performs a 5-step forward forecast. Use the model to make forecasts on a German monthly natural gas consumption dataset. Comparing the results of WOA-LASSO with other five other WOA-based hybrid models and Cross-Validation based models for prediction results, it is found that WOA-LASSO has the smallest MAPE in each step of the 5-step prediction, and the numerical results are between 8.273% and 9.867%. Moreover, when comparing WOA with the conventional optimization scheme Cross-Validation, it is found that WOA can obtain better model hyperparameters, which can effectively enhance the generalization performance and prediction accuracy of the model.

Keywords: Least Absolute Shrinkage and Select Operator; Whale Optimization Algorithm; Nonlinear Auto-Regressive; Natural Gas Consumption.

Introduction

In the current context of the Russia-Ukraine conflict, Germany is in a gas crisis after Russia stopped supplying gas to Germany via Nord Stream 1 (55.2% of Germany's gas imports), shaking the industrial and economic foundations of the EU's number one economy. The ability of the German energy regulator to accurately forecast gas consumption in the medium and long term is of great importance to the country's development. Therefore, a study of natural gas consumption is not only useful for companies to plan their investments, but also to ensure the healthy development of their natural gas industry and to ensure the construction of a modern and efficient energy system.

Currently, common prediction models are generally linear regression models [1], Exponential Smoothing Models [2], Neural network predictive model [3], System dynamic prediction model [4], Grey Models [5][6] etc. The swarm intelligence optimization algorithm is a class of optimization techniques based on iterative evolutionary search of populations, which is more suitable for handling and solving large-scale optimization problems due to its strong global search capability, potential parallelism and distributed nature [6][7]. Whale Optimization Algorithm (WOA) is a new intelligent optimization bionic algorithm newly proposed by Mirjalili[9]. It originates from the simulation of humpback whale group hunting behaviour in nature, and achieves the purpose of optimizing the search through the process of searching, encircling, pursuing and attacking the prey by the whale group. The WOA algorithm is effective for constrained optimization problems with non-uniformly sparse arrays and is similar to the particle swarm optimization (PSO) algorithm and the gravitational search algorithm (GSA). algorithm (GSA) algorithm [9],the WOA algorithm has greater advantages in terms of computational speed, solution accuracy and robustness.

In the face of the high-dimensional massive dataset of this paper on natural gas, the selection of the method for feature dimensionality reduction [10][11] is also particularly important. Traditional feature selection methods such as stepwise regression [12], optimal subset selection [13], ridge regression methods[14], and principal component regression can only achieve some of these objectives. In order to find a feasible solution to this problem, Tibshirani proposed in 1996 a method called Bridge Regression, inspired by Frank's Bridge Regression [14] and Bireman's Nonnegative Garrote [13], Tibshirani proposed a new variable selection method called LASSO [15]. The LASSO method uses the absolute value function of the model coefficients as a penalty to compress the model coefficients so that coefficients with smaller absolute values are automatically compressed to zero, thus enabling both the selection of significance variables and the estimation of the corresponding parameters. Compared with the traditional method of feature selection, the LASSO method overcomes the shortcomings of the traditional method in selecting models very well. Many scholars have conducted in-depth research on the effective algorithm of this method. shooting algorithm was proposed by Fu [16] in 1998, and Osborne et al. proposed the Tonglen algorithm based on the fact that the path of the solution of LASSO is progressively linear. the minimum angle regression algorithm proposed by Efron et al [17]. solved the computational problem of LASSO well, making the LASSO method as a simple and effective feature selection algorithm and become widely popular. This paper therefore adopts the combined WOA-LASSO model for forecasting natural gas consumption in Germany, and the numerical experimental results show that the combined model can obtain accurate and effective forecasting results.

2. Design of Combined Forecasting Model

In this section, the detailed mathematical model of the LASSO regression and WOA algorithm used in this paper will be given in **Section 2.1** and **Section 2.1**, respectively. And the complete multi-step forecasting model based on the NAR formulation will be given at **Section 2.3.**

2.1 LASSO Regression model

LASSO (Least Absolute Shrinkage and Select Operator) is a linear regression method that adopts $L1$ regularization. By using $L1$ regularization, the weight of some learned features will be zero, so as to achieve the purpose of sparse and feature selection. The general linear regression model can be expressed as:

$$y = X\beta + \varepsilon \tag{1}$$

where y is a $n \times p$ matrix, β is the positional parameter, and ε is a random error.

The basic idea of least square method is to make the ε as small as possible, the process of minimization problem is the following expression:

$$J(\beta) = \|\varepsilon\|^2 = \|y - X\beta\|^2 = (y - X\beta)^T(y - X\beta) \qquad (2)$$

by the optimality condition, set the partial derivative of $J(\beta)$ with respect to β to be 0. We can obtain the solution:

$$\hat{\beta} = (X^T X)^{-1} X^T y \qquad (3)$$

LASSO regression is to add a 1-norm to the minimized objective function $J(\beta)$, so the loss function of LASSO regression formed in the following form:

$$J(\beta) = \|y - X\beta\|_2^2 + \lambda\|\beta\|_1 \qquad (4)$$

where $\lambda \geq 0$. Same as above solution, we can obtain:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y \qquad (5)$$

It is easier to get sparse solutions in LASSO regression by drawing and comparing. Because the 1-norm contains some non-differentiable angle points on the axis. Obviously, LASSO is easier to get the 0 parameters of the model. Therefore, LASSO regression is good at feature selection, that is, removing irrelevant or redundant features.

2.2 The Whale Optimization Algorithm

The WOA algorithm is inspired the special hunting method of the humpback whales. The hunting behavior is divided into three stages: Encircling, Bubble-net attacking, and searching for prey. The specific mathematical modeling steps for these three stages are described below.

Encircling: The initial position of the prey is unknown. When the humpback whales find the prey, they will encircle them. It is assumed that the current solution is the position of the prey or close to the optimal position. After the best search agent is defined, the other search agents will thus try to update their position to the best search agent. This progress modeled as :

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A} \cdot \vec{D} \qquad (6)$$

$$\vec{D} = \left|\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)\right| \qquad (7)$$

where $\overrightarrow{X^*}(t)$ represents the best solution of the current iteration. $\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a}$ and $\vec{C} = 2 \cdot \vec{r}$ are coefficient vectors, $\vec{r}$ is a random vector between 0 and 1, $\vec{a}$ is linearly decreased from 2 to 0 over the iteration course.

Bubble-net attacking: The humpback whales attack prey in two ways: shrinking encircle and spiral update mechanism. Assume that there is a probability of 50% to choose the two way to update the position of whales. This progress modeled as follows:

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \vec{A} \cdot \vec{D} & if\ p < 0.5 \\ \overrightarrow{D'} \cdot e^{nl} \cdot \cos(2\pi l) + \overrightarrow{X^*}(t) & if\ 0.5 \leq p < 1 \end{cases} \qquad (8)$$

Where n is a constant, and l is a random value in $[0, 1]$.

searching for prey: The variation of $\vec{A}$ can be utilized to determine the current optimal solution. Actually, the humpback whales search randomly based on each other's positions. When $\vec{A} > 1$ or $\vec{A} < -1$, the current whale may far away from the reference whale. In contrast to the bubble-net attacking phase, update the position of a search agent according a randomly choose one instead of the current best search agent in searching phase, which can enhance the global search ability. The mathematical model is as follows:

$$\vec{D} = \left|\vec{C} \cdot \overrightarrow{X_r} - \vec{X}\right| \tag{9}$$

$$\vec{X}(t+1) = \overrightarrow{X_r} - \vec{A}.\vec{D} \tag{10}$$

where $\overrightarrow{X_r}$ is a random vector represent a random whale chosen from the current iteration.

2.3 Complete multi-step forecasting scheme based on WOA-LASSO

Hyperparameter optimization is the most significant problem in machine learning problems. In this paper, we will not use conventional *k-fold* cross-validation, but choose to use out-of-sample holdout validation, since the use of *k-fold* cross-validation in time series models is somewhat controversial.

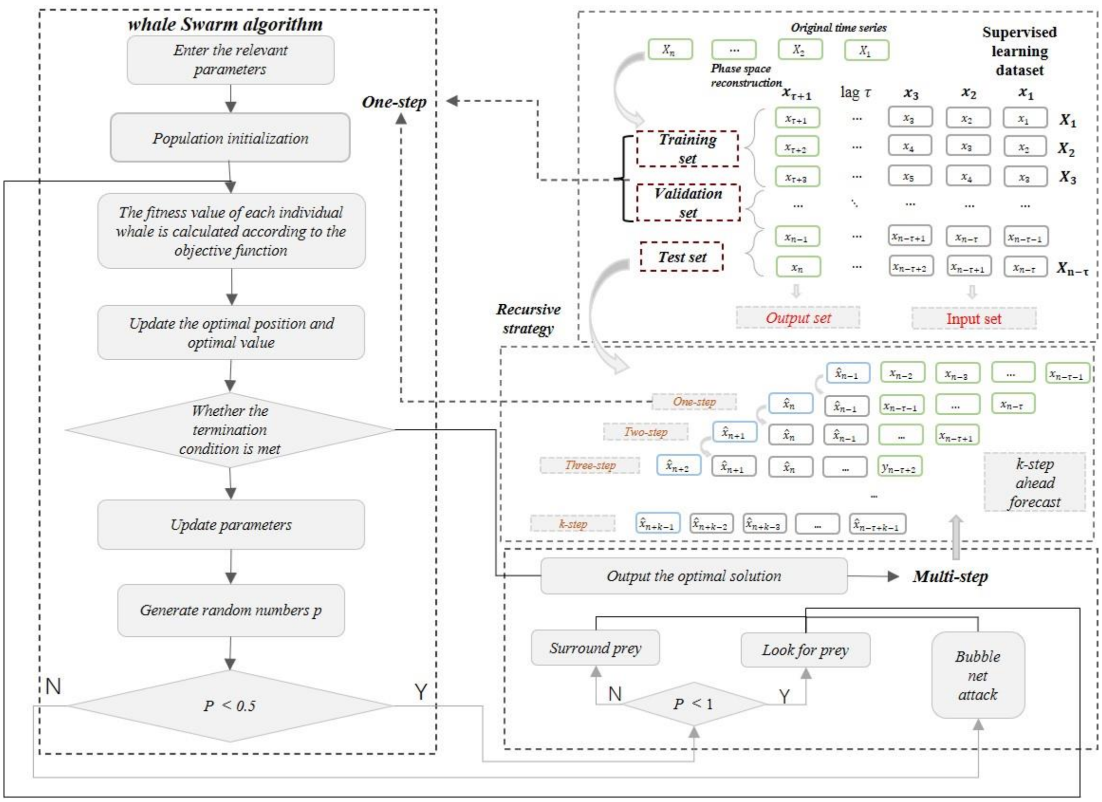

Figure 1 Complete Algorithm Process

Nonlinear Auto-Regressive (NAR) model based on the concept of phase space reconstruction [8], is used to reconstruct the dataset. Given a univariate time series $X = \{x_1, x_2, x_3, \dots, x_n\}$, choose a τ value as the time lag, this time series is transformed into a new dataset Φ that can be used for supervised learning, which form can be expressed as follows:

$$\Phi = \begin{pmatrix} x_1 & x_2 & \cdots & x_\tau & x_{\tau+1} \\ x_2 & x_3 & \cdots & x_{\tau+1} & x_{\tau+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n-\tau} & x_{n-\tau+1} & \cdots & x_{n-1} & x_n \end{pmatrix} \tag{11}$$

Divide the new dataset according to the ratio of about *8:1:1*, and the divided datasets are respectively training set, validation set and test set. First use the WOA algorithm to find the optimal hyperparameters of the model on the training set, then use the validation set for validate the hyperparameters, MAPE is the smallest on the validation set as the goal of optimization.

$$MAPE = min \frac{1}{n} \sum \left| \frac{x_j - \widehat{x_j}}{x_j} \right| \times 100\% \tag{12}$$

and finally use the test set for multi-step forecasting. The complete multi-step forecasting strategy is shown in **Figure 1**.

3. Dataset Description

In our research, the data used is from the publicly available German natural gas consumption (NGC) dataset in the Eurostat (https://ec.europa.eu/eurostat), which collects monthly NGC data from Jan 2014 to May 2022 for a total of 100 months. The first 80 points are used to train the model, 81-90 points are used to validate the model, and 91-100 points are used for multi-step forecasting. The dataset is shown in **Figure 2**.

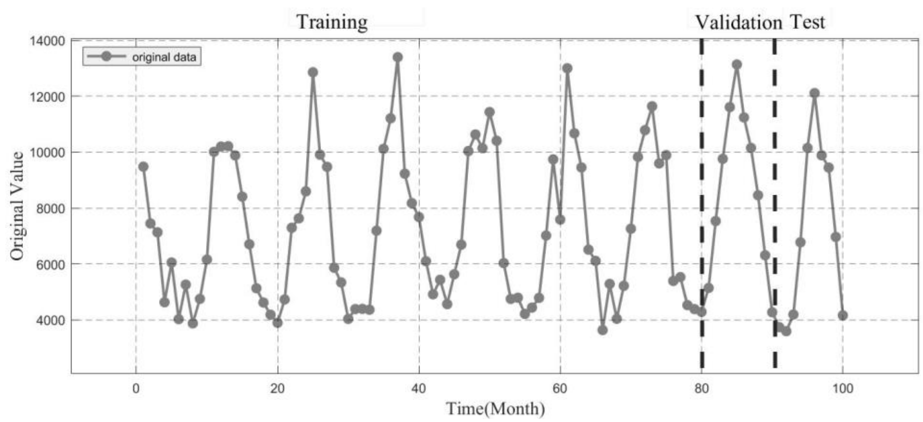

Figure 2 Raw data on natural gas consumption in Germany

4. Multi-step forecasting results and Discussion

In this Section, comparing the forecasting results of WOA-LASSO with five other classical machine learning models (including SVR with RBF kernel, Random Forest, MLP, LSVR, XGBoost) optimized with WOA, and compared with the forecasting results of the six models including LASSO using grid search cross validation method, which aims to verify the proposed optimization scheme is superior than the conventional optimization scheme.

4.1 Analysis of the forecasting results

In order to quantitatively analyze the performance of the proposed model, the MAPE in the multi-step comparison is used as a comparison metrics, and the smaller the metrics, the better the performance of the model. In all experiments, a time lag of 5 was chosen to ensure that as much data as possible is used for validation and multi-step forecasting process, each model predicts 5 steps forward.

Table 1 MAPE(%) of the forecasting models

optimizer	steps	LASSO	SVR	RF	MLP	LSVR	XGBoost
WOA	**1-step**	***8.273***	10.120	15.673	17.018	13.284	13.866
	2-step	***9.414***	12.483	13.321	24.515	14.107	18.247
	3-step	***9.085***	12.806	13.135	29.351	13.484	24.110
	4-step	***9.867***	13.429	14.868	27.401	12.869	23.095
	5-step	***8.774***	11.595	11.907	21.887	17.123	20.422
CV	**1-step**	9.728	10.528	15.279	40.590	13.284	14.064
	1-step	11.233	12.687	12.791	52.465	14.107	17.150
	3-step	9.750	12.843	11.795	59.626	13.484	18.469
	4-step	10.673	13.559	12.940	68.384	12.869	19.818
	5-step	9.385	11.743	11.300	68.449	17.123	17.418

Applying the proposed model to the forecast of monthly natural gas consumption in Germany. The models for comparison are the classic SVR and LSVR models, the RF and XGBoost models with super generalization capabilities, and the neural network model MLP, a total of three types.

It can be clearly seen that from **Table 1**, the MAPE of the proposed WOA-LASSO combined forecasting model is from 8.273% to 9.867%. The MAPE of each step is lower than other combined models, and lower than CV-LASSO , which shows that the optimization scheme used in this paper is better than the conventional optimization scheme. The prediction data of each model on the test set is shown in **Figure 3**.

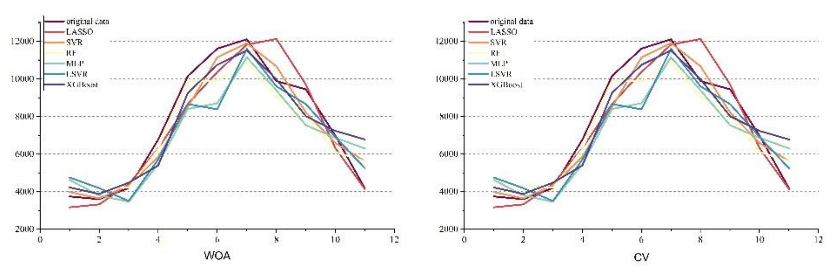

Figure 3 Predict Data in the test set

4.2 Discussion

In this section, we discuss the improvement of the model performance by the optimization algorithm. Since the WOA algorithm was proposed, it has been widely used to solve various complex nonlinear optimization problems due to its few adjustment parameters and the advantages of being easy to jump out of local convergence.

WOA-LASSO and CV-LASSO are presented in **Figure 4**, It can be clearly seen that after using WOA for hyperparameter optimization, the performance of the LASSO model has been significantly improved at each step. The biggest improvement is the first step and the second step, which are 1.498% and 1.824% respectively.

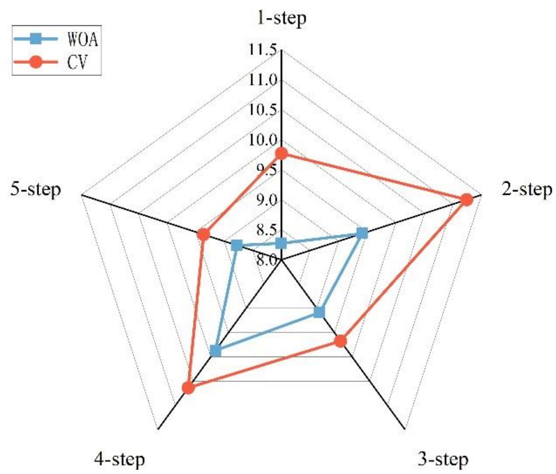

Figure 4 MAPE results of LASSO by using WOA and CV

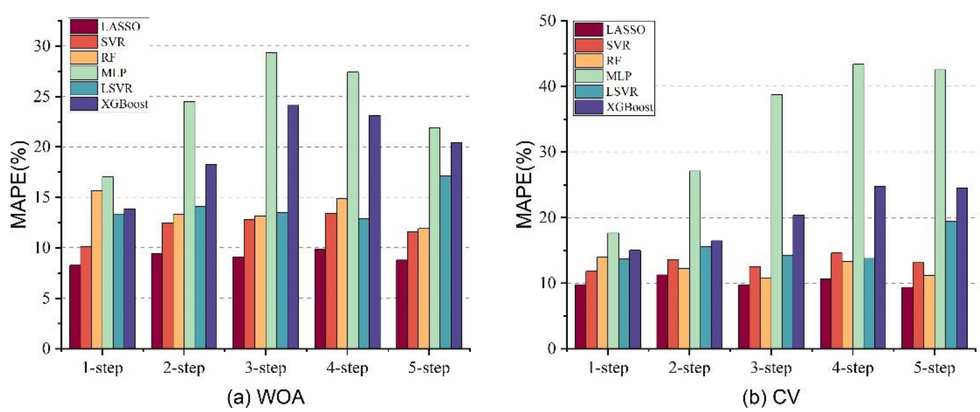

Figure 5 MAPE results of all models by using WOA and CV

And all other models are presented in **Figure 5**. Compared with the traditional CV optimization algorithm, after using WOA, except for XGBoost, the predictive ability of the other five models is enhanced to a certain extent. Especially the MLP model, after using WOA, the MAPE in the first step is reduced by 23.572%, and the fifth step is the most reduced by 46.562%.

In summary, WOA has the characteristics of fast convergence, and it is easy to jump out of the local optimum globally to achieve global convergence, can effectively improve the generalization performance of machine learning models.

5. Conclusion

In the context of the current energy shortage, in order to better predict the consumption of natural gas, relevant decision makers can plan in advance and use it efficiently. The WOA-LASSO combined forecasting model based on the NAR formula is used for multi-step forecasting of natural gas consumption in this paper, and the model is applied to the forecasting of monthly natural gas consumption in Germany. Comparing this model with SVR, LSVR, RF, XGBoost, and MLP five combined models based on WOA, it is found that its prediction accuracy is the highest. The MAPE of each step of its *5-step* forecast is from 8.273% to 9.867%. It also compared WOA with the conventional optimization scheme cv, and found that WOA can better optimize the model parameters, so that the model can obtain stronger generalization performance and prediction accuracy.

References

[1] Wang Huiwen, Meng Jie. A predictive modeling approach to multiple linear regression [D]. , 2007.
[2] Yin Guangzhi, Zhang Weizhong, Zhang Dongming, et al. Landslide prediction based on the combination of exponential smoothing and regression analysis[J]. Geotechnics, 2007, 28(8): 1725-1728.
[3] Chen C-Y, Wang S-J, Shen S-K. Artificial neural network prediction model for stability of slope rock[J]. Journal of Geotechnical Engineering, 2001, 23(2): 157-161.
[4] Liu JL, Zang JH, He YQ. System dynamics model for urban water demand prediction[J]. China Water Supply and Drainage, 2005, 21(6): 31-34.
[5] Sun Chenjun, Wang Cuiru, Zhang Jiangwei, et al. Improvement and application of residual gray prediction model[J]. Statistics and Decision Making, 2005: 19-20.
[6] Zhou Yin. The application of residual grey forecasting model in logistics demand forecasting [J]. Railway Transport and Economy, 2007, 29(11): 59-61.
[7] Huang Guangqiu, Li Tao, Lu Qiuqin. Artificial memory optimization algorithm[J]. Systems Engineering Theory and Practice, 2014, 34(11): 2900-2912.
[8] Sivakumar B, Jayawardena A W, Fernando T. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches[J]. Journal of hydrology, 2002, 265(1-4): 225-245.
[9] Mirjalili S, Lewis A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51–67.
[10] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006, 68(1): 49-67.
[11] Guyon I, Elisseeff A. An introduction to variable and feature selection[J]. Journal of machine learning research, 2003, 3(Mar): 1157-1182.
[12] Friedman J H. Multivariate adaptive regression splines[J]. The annals of statistics, 1991, 19(1): 1-67.
[13] John G H, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem[M]//Machine learning proceedings 1994. Morgan Kaufmann, 1994: 121-129.
[14] Hastie T, Taylor J, Tibshirani R, et al. Forward stagewise regression and the monotone LASSO [J]. Electronic Journal of Statistics, 2007, 1: 1-29.
[15] Torgo L, Gama J. Regression by classification[C]//Brazilian symposium on artificial intelligence. Springer, Berlin, Heidelberg, 1996: 51-60.

[16] Yang G. Improved Shooting Algorithm Based on Zone and Tangential Arc in Robot Soccer[C]//Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA09). 267-269.

[17] Efron B, Hastie T, Johnstone I, et al. Least angle regression[J]. The Annals of statistics, 2004, 32(2): 407-499.

[18] Sivakumar B, Jayawardena A W, Fernando T. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches[J]. Journal of hydrology, 2002, 265(1-4): 225-245.