

Research Article

Soft Sensor Modeling Method Based on SPA-GWO-SVR for Marine Protease Fermentation Process

Zhu Li,¹ Khalil Ur Rehman ,¹ Liu Wenhui,¹ and Faiza Atique²

¹School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China

²Department of Physiology, GC University Faisalabad (GCUF), Faisalabad, Pakistan

Correspondence should be addressed to Khalil Ur Rehman; 5102180315@stmail.ujs.edu.cn

Received 31 December 2020; Revised 22 February 2021; Accepted 1 March 2021; Published 20 March 2021

Academic Editor: Yongji Wang

Copyright © 2021 Zhu Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The marine protease fermentation process is a highly nonlinear, time-varying, multivariable, and strongly coupled complex biochemical reaction process. Due to the growth and reproduction of living organisms, the internal mechanism is very complicated. Some key variables (such as cell concentration, substrate concentration, and enzyme activity) that directly reflect the fermentation process's quality are difficult to measure in real-time by traditional measurement methods. A soft sensor model based on a support vector regression (SVR) is proposed in this paper to resolve this problem. To further improve the model's prediction accuracy, the grey wolf optimization (GWO) algorithm is used to optimize the critical parameters (kernel function width σ , penalty factor c , and insensitivity coefficient ϵ) of the SVR model. To study the influence of selecting auxiliary variables on soft sensor modeling, the successive projection algorithm (SPA) is used to determine the characteristic variables and compare them with grey relation analysis (GRA) algorithm. Finally, the Excel spreadsheet data was called by MATLAB programming, and the established SPA-GWO-SVR soft sensor model predicted crucial biological variables. The simulation results show that the SPA-GWO-SVR model has higher prediction accuracy and generalization ability than the traditional SPA-SVR model. The real-time monitoring was processed by MATLAB software for the marine protease fermentation process, which met the requirements of optimal control of the marine protease fermentation process.

1. Introduction

In recent years, marine biotechnology's rapid development makes marine microbiological engineering occupy an essential proportion in the national economic system. Marine low-temperature alkaline protease is a new kind of protease source and, compared with other terrestrial proteases, can effectively solve such problems as high requirement of enzyme activity, instability, and inactivation in harsh environment in industrial production. These excellent characteristics make it widely used in the washing industry, environmental protection, food processing, and national defence [1, 2]. In the actual fermentation process, to increase the production efficiency and product quality of enzyme preparation and reduce the economic cost, the fermentation process's environmental variables should be controlled in a specific range. However, the process of cell growth,

reproduction, and metabolic enzyme production is significantly reflected by the external environment. There is a complicated non-linear dynamic relationship among various state variables of the fermentation process, which is difficult to decouple. The critical biological variables reflecting the quality of fermentation have severe defects in measurement stability and price. In the fermentation process, temperature, airflow, pH value, dissolved oxygen concentration, motor stirring speed, etc., can be detected by mature hardware sensors. However, cell concentration, substrate concentration, and enzyme activity, which are target quality variables that play a significant role in the fermentation process, are challenging to measure in real-time. Therefore, studying the real-time measurement method of the fermentation process's key variables is of great significance for improving the quality of fermentation products, reducing the consumption of raw materials, and power energy [3].

The traditional analytical mathematical model cannot meet the requirements of fermentation process control. Using soft sensor modeling to establish a mathematical model between auxiliary variables (easy to measure) and target quality variables (difficult to measure) in real-time provides a new way for real-time and accurate measurement of biological fermentation variables [4]. The model structure of the soft sensor is shown in Figure 1. This figure shows the basic principle of a soft sensor model where inputs from a bioprocess are measured by hardware sensor/sensors, and new information is estimated. Soft sensor modeling can overcome many shortcomings of manual analysis and real-time instrumentation and has become the leading direction of the development of advanced control technology in the fermentation industry [5]. The core of soft sensor modeling is to establish a mathematical model. The accuracy and generalization of the model directly determine the results of soft measurement. The main methods of soft sensor modeling include process mechanism analysis, state estimation [6], regression analysis [7], artificial neural network (ANN) [8], and deep neural networks (DNN) [9], etc.

However, it is challenging to obtain the system state-space model by relying entirely on mechanism analysis to establish an accurate soft sensor model. Regression analysis requires a lot of data and is more sensitive to measurement errors. The prediction method based on neural networks has become a research hotspot in the field of soft-sensing. However, for a very complex non-linear system such as the fermentation process of marine microorganisms, if there is no prior knowledge, the neural network method is blindly applied, and the measurement of critical biological variables cannot be solved well. In the literature [10], the back-propagation neural network (BPNN) is used to measure the flocculent bacteria's fermentation yield and flocculation rate. However, the BPNN has the disadvantages of slow convergence and weak anti-interference ability. Deep neural networks have also been employed for process soft sensor modeling in the recent few years. Deep learning-based model is suitable for large training samples, and it enjoys high efficiency and robustness. In the literature [9], a layer-wise data augmentation strategy with stacked auto-encoder (LWDA-SAE) is applied to predict the boiling points in a hydrocracking process. In the literature [11], a dynamic convolutional neural network (DCNN) strategy is introduced to learn hierarchical local non-linear dynamic features for soft sensor modeling. In the literature [12], a spatiotemporal attention-based long short-term memory (LSTM) network is applied to predict an industrial process's quality variables. In another paper, the researchers applied the DNN-based model to predict the quality variables of fermentation process [13].

In this article, the support vector machine regression (SVR) algorithm is applied to predict the quality variables of the marine protease fermentation process, a common and useful method used for small amounts of data. It is a machine learning method based on statistical learning theory and structural risk minimization [14]. However, in SVR modeling, normalization parameters and kernel parameters

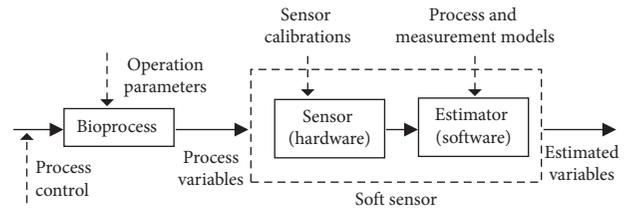


FIGURE 1: Principle of the soft sensor model.

are the parameters that must be optimized, and their values will directly affect the training and generalization performance of the model. Commonly used parameter optimization methods include grid search with cross-validation (GSCV) and genetic algorithms (GA). The GSCV method and GA algorithm are computationally intensive and may fall into the local optimum [15]. Therefore, a grey wolf optimization (GWO) algorithm is used to optimize the SVR soft sensor model's essential parameters. Besides, in soft sensor modeling, the choice of auxiliary variables is also necessary. Many scholars have ignored the influence of selecting additional variables on the soft sensor modeling of the microbial fermentation process. If some unimportant independent variables are introduced into the soft sensor model, the accuracy of the model will be reduced. The successive projection algorithm (SPA) is considered to be one of the best indicators for evaluating the correlation of variables in the process of fermentation and compared with the grey relation analysis (GRA) algorithm [16]. Finally, the proposed soft sensor modeling method based on SPA-GWO-SVR is applied to predict the target quality variables such as cell concentration, substrate concentration, and the marine protease fermentation process's enzyme activity. The simulation results show that the constructed soft sensor model has a simple structure, high modeling accuracy, and reasonable practicability.

2. Materials and Methods

2.1. Successive Projection Algorithm (SPA). SPA [17] is an influential variable selection method, which can minimize the correlation between auxiliary variables and eliminate redundant variables. Suppose $X_{M \times J}$ is the original data matrix, where M is the total number of samples, J is the number of auxiliary variables, N is the number of selected characteristic variables, and $X_{k(0)}$ is the initial variable. Start with the first auxiliary variable, calculate its projection on the remaining variables, introduce the variable with the largest projection vector into the set, and end after N iterations. The specific implementation steps are as follows:

Step 1. Select a column of data variable j , assign it to X_j , and write it as $X_{k(0)}$.

Step 2. Let S be the set of remaining variables, expressed explicitly as $S = \{j | j \notin [k(0), k(1), \dots, k(N-1)], 1 \leq j \leq J\}$.

Step 3. Calculate the projection of the remaining variables according to formula (1).

$$P_{X_j} = X_j - (X_j^T X_{k(n-1)}) X_{k(n-1)} (X_{k(n-1)}^T X_{k(n-1)})^{-1}. \quad (1)$$

Step 4. Select the maximum projection value according to formula (2):

$$k(n) = \arg(\max(\|P_{X_j}\|)), \quad j \in S. \quad (2)$$

Step 5. Let $X_j = P_{X_j}$, $j \in S$ use the maximum projection value in Step 4 as the comparison value of the next iteration.

Step 6. Let $n = n + 1$, if $n < N$, return to Step 2 to continue the iteration.

Step 7. Calculate the root mean square error (RMSE) corresponding to each variable subset by combining the multivariate correction method and the cross-validation method, select the variable subset with the smallest RMSE, determine the set with the smallest number of variables, and use the variables as characteristic variables.

2.2. Support Vector Regression (SVR) Algorithm. SVR is a regression prediction algorithm suitable for dealing with non-linear problems. The algorithm mainly uses ascending dimensions to construct a linear decision function in a high-dimensional space to achieve linear regression [18]. SVR uses a similar theory as the SVM for classification problems, with just a few slight differences. The machine-like SVM can be equivalent to the network shown in Figure 2. This figure shows the working procedure of SVR, where x^m is the input vector, $k(x_i, x_j)$ is a kernel function, and a_i is weights. Suppose the training sample set is $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, $x_i, y_i \in R$ an auxiliary variable, and y_i the soft sensor model's critical biological variables. The purpose is to find a regression model $y = w \cdot x + b$ on a high-dimensional space. To ensure the flatness of the model curve, we need to find the smallest w and finally transform the optimization problem into

$$\min R(w, w) = \frac{1}{2} (w, w) + c \sum_{i=1}^l (\xi_i + \xi_i^*), \quad (3)$$

$$s.t. (y - w \cdot x - b) \in [-\varepsilon - \xi_i, \varepsilon + \xi_i].$$

C is a regulation parameter, w is a weight vector, b is bias unit, ξ and ξ^* are slack variables, and ε is the insensitivity coefficient. By introducing Lagrange multipliers a_i and a_i^* , the Lagrange function is constructed, and formula (3) is further transformed into a dual problem:

$$\begin{aligned} \max \varphi(a) = & \sum_{i=1}^l (a_i - a_i^*) y_i + \varepsilon \sum_{i=1}^l (a_i - a_i^*) \\ & - \frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*) (a_j - a_j^*) (x_i \cdot y_j), \end{aligned} \quad (4)$$

$$s.t. \begin{cases} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ a_i, a_i^* \in [0, c] \end{cases}$$

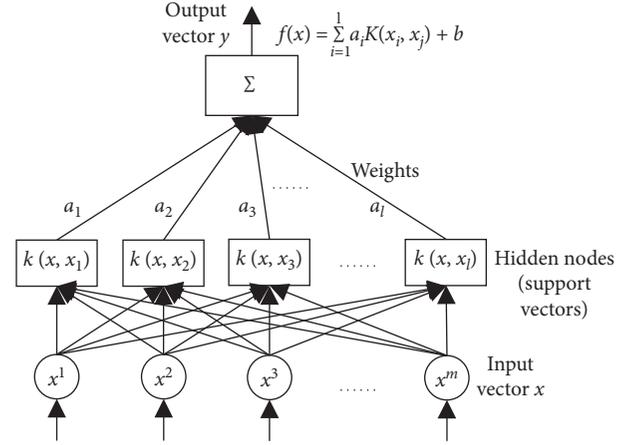


FIGURE 2: Support vector regression architecture.

To simplify the calculation, the kernel function is introduced to replace the inner product calculation. This paper uses the radial basis kernel function (RBF) with fewer parameters:

$$k(x_i, x_j) = \exp \left\{ -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right\}. \quad (5)$$

Among them, $k(x_i, x_j)$ is a kernel function, and σ is the width parameter of the kernel function. By seeking partial derivatives, solve for a_i, a_i^* and b , where the corresponding vector for $a_i \neq 0$ or $a_i^* \neq 0$ is the support vector. The final analytical expression of SVR is

$$y = \sum_{i=1}^l (a_i - a_i^*) k(x_i, x_j) + b. \quad (6)$$

Although SVR has good generalization ability and good regression prediction effect for small samples and non-linear data, the critical parameters in the SVR model, such as the kernel function width parameter σ , penalty factor c , and insensitivity coefficient ε , and the quality of the initial parameter settings, will have a significant impact on the stability and prediction accuracy of the model. Therefore, to solve this problem, this work uses an intelligent optimization algorithm based on GWO to initialize and optimize the parameters of the SVR soft sensor model of the marine protease fermentation process. The GWO algorithm [19] proposed in 2014 is significantly better than the GA algorithm, particle swarm optimization algorithm (PSO), and optimization effect.

2.3. GWO-SVR Prediction Model. Grey wolf optimization (GWO) algorithm is a new type of swarm intelligence optimization algorithm, by simulating the leadership mechanism at all levels of the grey wolf population and the process of wolves encircling, hunting, and attacking to achieve the purpose of optimization search [20].

First, calculate each wolf's fitness in the wolf pack and arrange them in order from high to low. The wolf pack is

divided into α wolf, β wolf, δ wolf, ω wolf, among which α wolf, β wolf, and δ wolf are looking to play a leading role in the optimization process. When the grey wolf surrounds the prey, the position is updated by formulas (9) and (10):

$$\begin{aligned} D &= |C \cdot X_p(t) - X(t)|, \\ X(t+1) &= X_p(t) - A \cdot D, \\ st &\begin{cases} A = 2a \cdot r_1 - a; \\ C = 2 \cdot r_2. \end{cases} \end{aligned} \quad (7)$$

Among them, t represents the number of iterations; X_p represents the position of the prey, which is the optimal global solution; X represents the position of the current wolf pack; D represents the distance between the prey and the wolf pack; r_1 and r_2 represent a random value within $[0, 1]$ number; and a represents a number that drops from 2 to 0.

After the prey is surrounded, the wolves need to hunt, α wolf, β wolf, δ wolf are assumed to be the three global optimal solutions, and ω wolf updates the position again according to these three optimal solutions:

$$\begin{aligned} D_\alpha &= |C_1 \cdot X_\alpha - X|, \\ D_\beta &= |C_2 \cdot X_\beta - X|, \\ D_\delta &= |C_3 \cdot X_\delta - X|. \end{aligned} \quad (8)$$

Among them, X_α , X_β , and X_δ represent the positions of α wolf, β wolf, and δ wolf, respectively; C_1 , C_2 , C_3 represent random vectors; and X represents the position of the current solution; when the approximate distance between the current solution and the α wolf, β wolf, δ wolf, and ω wolf is calculated, the position $X(t+1)$ of the wolf that needs to be updated is as follows:

$$X(t+1) = \frac{X_\alpha + X_\beta + X_\delta - (A_1 \cdot D_\alpha + A_2 \cdot D_\beta + A_3 \cdot D_\delta)}{3}. \quad (9)$$

Among them, A_1 , A_2 , and A_3 represent random vectors. Finally, the wolves attack the prey, which is to determine the optimal solution. It is mainly by observing the value A to determine whether the optimal solution is found. If $|A| > 1$, it means that the wolf pack is getting farther and farther away from its prey, and the optimal solution cannot be concentrated; if it means that the wolf pack is concentrated on the prey, it can be captured.

Using GWO to optimize SVR is actually to find the optimal kernel function width parameter σ , penalty factor c , and insensitivity coefficient ε in the SVR through the above update process. The specific implementation steps are as follows:

Step 1. Initialize the parameters. Determine the training set and prediction set samples of the SVR model. Set the maximum number of iterations T , the population size N , and the optimization range of σ , c , and ε .

Step 2. Randomly generate grey wolf packs. The individual position vector of each grey wolf pack consists of σ , c and ε , and calculate the fitness value of each grey wolf.

Step 3. Classify the grey wolf population according to the fitness value, and update the individual position of the grey wolf population.

Step 4. Calculate each grey wolf individual's fitness value at the new position and compare it with the optimal fitness value of the previous iteration. If it is better than the fitness value, the grey wolf individual replaces the group's optimal fitness value. And replace the fitness value with the grey wolf individual position; otherwise, keep the original optimal fitness value.

Step 5. If the number of iterations exceeds the maximum number of iterations, the learning process ends, and the output global optimal position is the optimal value of σ , c , and ε in SVR; otherwise, return to Step 3 to continue parameter optimization.

Step 6. Use the optimized σ , c , and ε to establish the SVR model.

3. Results and Discussion

3.1. Fermentation Experiment. The experimental data have been collected from the fermentation control system platform, made by Jiangsu University, China. The fermentation tank was the RT_100L_Y model of Zhenjiang Ritai Biological equipment company. Marine protease fermentation is taken as an example for experimental verification, and the process of fermentation is conducted in the biological fermentation tank. The basic flow of the fermentation process is shown in Figure 3. To make the experiment close to the actual production process, the experiment process was designed as follows.

In the process of marine protease fermentation, the temperature of the fermentation tank is controlled at 25°C~28°C, the pressure of the fermentation tank is controlled at 0.03~0.05 Mpa, the dissolved oxygen is maintained at 35%~75%, the airflow is owned at 1000 L/h, and the speed is held at 300~500 r/min. The experiment sampling period is 15 minutes, and 11 auxiliary variables are acquired through various sensors and transmitted to the computer by a distributed control system (DCS) form a database. Under normal fermentation conditions, take samples every hour to analyze the cell concentration X , substrate concentration S , and enzyme activity P offline. X is determined by the mycelial dry weight method; S is determined by Fehling reagent method; P is determined by Folin-phenol reagent method. A total of 10 batches of fermentation data were obtained with 80 h span between every batch. The first 9 batches of fermentation data were used to train the soft sensor model for minimum error. And the 10th batch of fermentation data is used to test the identification accuracy of the model. In this research, the soft sensor model's training procedure was conducted on a laptop, with Intel Core i3-3110M CPU @2.40 GHz, 8 GB RAM. The software tools are Windows 10 pro, MATLAB R2018b, MathType, and Visio 2016. Necessary steps such as auxiliary variables selection, data pre-processing, hyper-parameter tuning, deciding kernel function, and establishment of the model, etc. have been

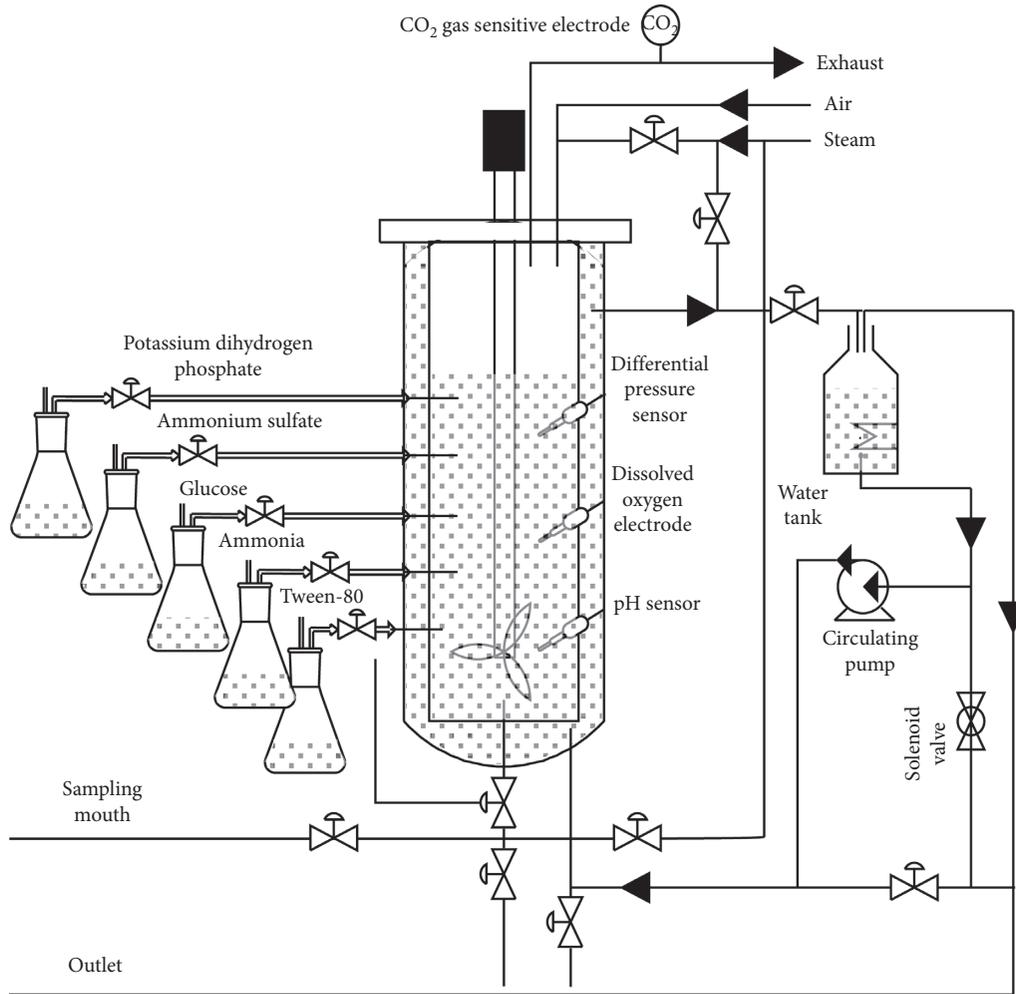


FIGURE 3: Technological process of marine protease fermentation process.

implemented for the development of SPA-GWO-SVR soft sensor model.

Data preprocessing performs a very significant role in various machine learning algorithms [21, 22]. In practice, several approaches work best after data normalization and whitening. Due to the yield environment's effect, the collection of a dataset is often subject to a variety of on-site noise interference, such as human-made noise or machine operating noise. After deciding the auxiliary variables and target quality variables of the soft sensor model, sufficient data samples are collected for each system variable. Still, the collected data samples comprise several errors and need to process the data. In this paper, the preferred processing techniques are mainly data filtering and data normalization. The normalization method is shown as follows:

$$i' = \frac{(i_{\max} - i)}{(i_{\max} - i_{\min})}, \quad (10)$$

where i' is the normalized data, i is the original sample data, i_{\max} is the maximum value of the sample data, and i_{\min} is the minimum value of the sample data.

3.2. Selection of Auxiliary Variables. We know that the marine protease fermentation is a complex non-linear time-varying process. During the process of fermentation, there are various types of internal and external environmental variables such as temperature t , tank pressure p , dissolved oxygen concentration DO, pH value, air flow l , motor speed r , volume v , acceleration rate of acid flow μ , alkali flow rate η , CO_2 concentration c_1 , and O_2 concentration c_2 that directly reflects the fermentation quality. To eliminate the influence of unnecessary auxiliary variables on model prediction accuracy, reduce the complexity, and increase the soft sensor model's calculation speed, this paper uses SPA and GRA algorithms to determine the input and output variables of the soft sensor model.

Using SPA to calculate the correlation between each auxiliary and critical variable, the results are shown in Table 1. The ones with the correlation degree exceeding 0.7 are selected as the auxiliary feature variables. To ensure the model's performance, the range of the number of variables selected by SPA is set to be 3–10, and the optimal number of characteristic variables is determined according to the RMSE under different numbers of variables. Figure 4 is a trend

TABLE 1: Correlation between auxiliary variables and key biological variables.

Auxiliary variables	Correlation degree		
	Cell concentration	Matrix concentration	Enzyme activity
Temperature t	0.2457	0.2469	0.4236
Tank pressure p	0.7164	0.7051	0.5862
Dissolved oxygen concentration	0.8853	0.8452	0.9645
pH value	0.8729	0.9124	0.8166
Airflow l	0.5321	0.4712	0.7033
Motor speed r	0.1983	0.2210	0.1875
Tank volume v	0.7088	0.6820	0.7040
Acceleration rate of acid flow μ	0.9441	0.9864	0.8769
Alkali flow rate η	0.9087	0.9753	0.8934
CO ₂ concentration CO _{2.con}	0.8865	0.7922	0.7965
O ₂ concentration O _{2.con}	0.7230	0.7894	0.7847

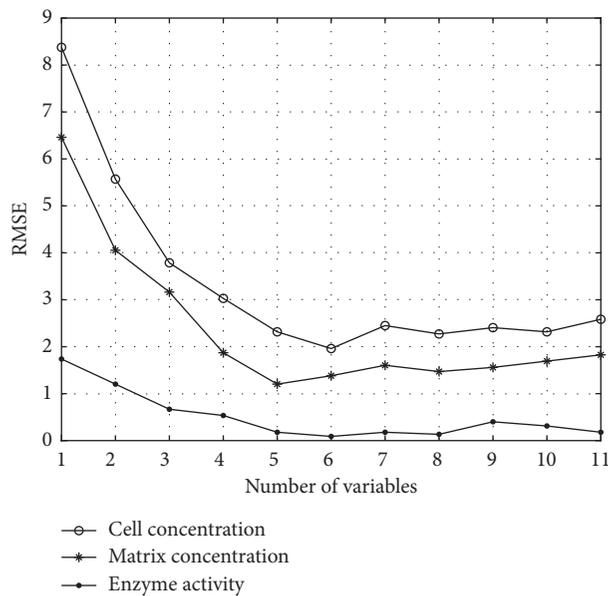


FIGURE 4: The influence of the number of variables selected by SPA algorithm.

TABLE 2: Selection of auxiliary variables based on GRA and SPA algorithm.

Key variables	Variable selection method	Number of variables	Auxiliary variables
Cell concentration (X)	Full	11	$t, p, DO, pH, l, r, v, \mu, \eta, c_1, c_2$
	GRA	8	$p, DO, pH, v, \mu, \eta, c_1, c_2$
	SPA	6	$DO, pH, v, \mu, \eta, c_1$
Matrix concentration (S)	Full	11	$t, p, DO, pH, l, r, v, \mu, \eta, c_1, c_2$
	GRA	7	$p, DO, pH, \mu, \eta, c_1, c_2$
	SPA	5	DO, pH, μ, η, c_1
Enzyme activity (P)	Full	11	$t, p, DO, pH, l, r, v, \mu, \eta, c_1, c_2$
	GRA	8	$DO, pH, l, v, \mu, \eta, c_1, c_2$
	SPA	6	$DO, pH, l, \mu, \eta, c_1$

diagram of the RMSE values of the three key variables changing with the number of selected variables. It can be found that, with the increase of the number of selected variables, the overall RMSE shows a downward trend. When the number of variables reaches 6, 5, and 6, the change trend

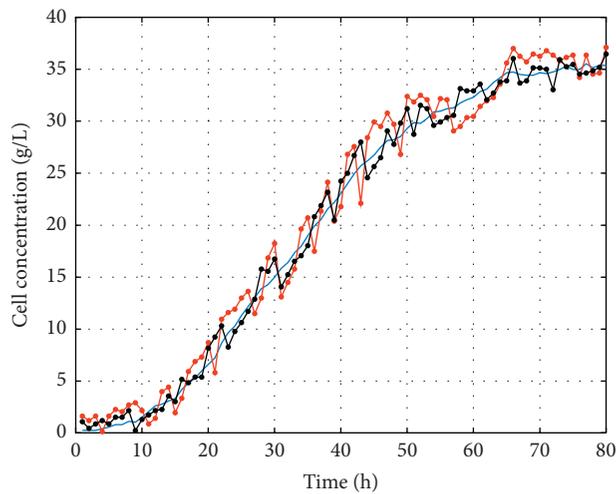
of the RMSE value is gentle. Since the larger number of auxiliary variables will lead to the higher complexity of the model, the optimal number of variables is finally determined to be 6, 5, and 6. The feature auxiliary variables selected based on GRA and SPA are shown in Table 2.

TABLE 3: Comparison of different variable selection methods.

Key variables	Variable selection method	Training set		Test set	
		R^2	RMSE	R^2	RMSE
X	Full	0.9623	2.4120	0.9672	2.5362
	GRA	0.9742	1.9665	0.9726	1.9965
	SPA	0.9851	1.9057	0.9832	1.9430
S	Full	0.9664	1.8402	0.9623	1.8635
	GRA	0.9690	1.1521	0.9687	1.1854
	SPA	0.9798	1.1422	0.9772	1.1693
P	Full	0.9682	0.1701	0.9678	0.1726
	GRA	0.9734	0.0852	0.9705	0.0921
	SPA	0.9845	0.0389	0.9828	0.0401

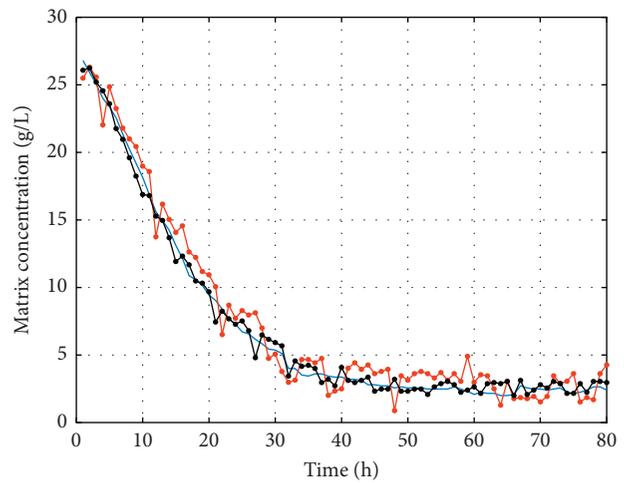
TABLE 4: Quality prediction results based on SPA-GWO-SVR soft sensor model.

Key variables	Variable selection method	Training set		Test set	
		R^2	RMSE	R^2	RMSE
X	Full	0.9783	1.8196	0.9763	1.8254
	GRA	0.9870	1.0641	0.9841	1.0897
	SPA	0.9945	1.0368	0.9933	1.0583
S	Full	0.9707	1.0026	0.9698	1.0254
	GRA	0.9811	0.6324	0.9786	0.6580
	SPA	0.9952	0.5066	0.9940	0.5561
P	Full	0.9794	0.0811	0.9781	0.0864
	GRA	0.9810	0.0301	0.9812	0.0348
	SPA	0.9982	0.0156	0.9971	0.0164



— Actual value
 —●— SPA-SVR
 —●— SPA-GWO-SVR

(a)



— Actual value
 —●— SPA-SVR
 —●— SPA-GWO-SVR

(b)

FIGURE 5: Continued.

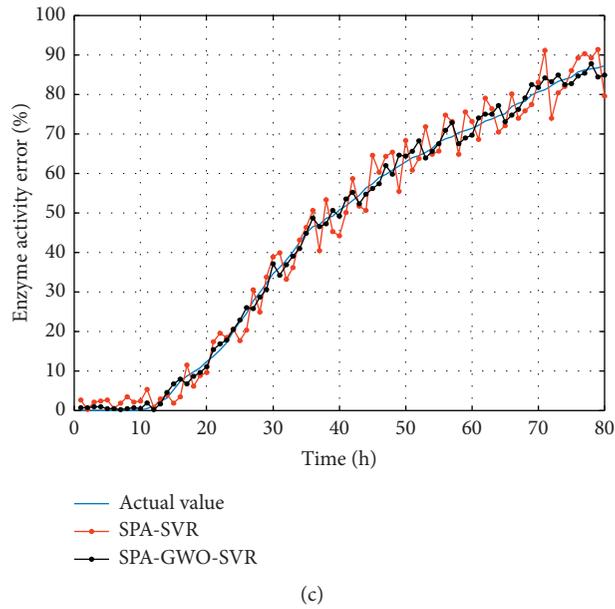


FIGURE 5: Comparisons of the predicted result with the reference of actual value for the process of marine protease fermentation. (a) Prediction results of cell concentration. (b) Prediction results of matrix concentration. (c) Prediction results of enzyme activity.

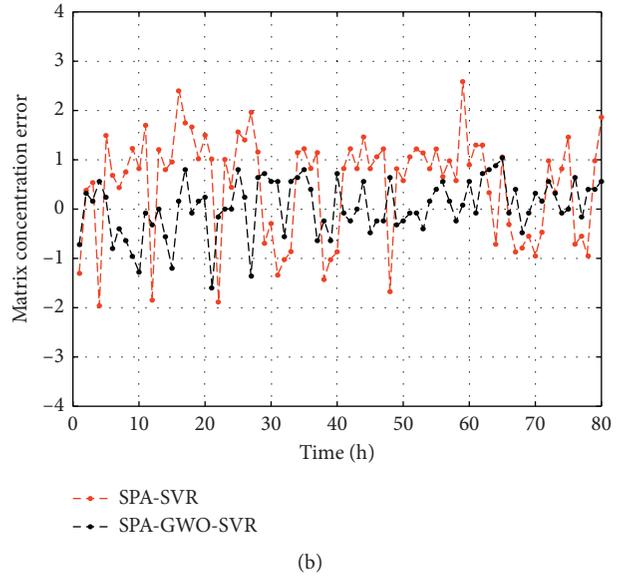
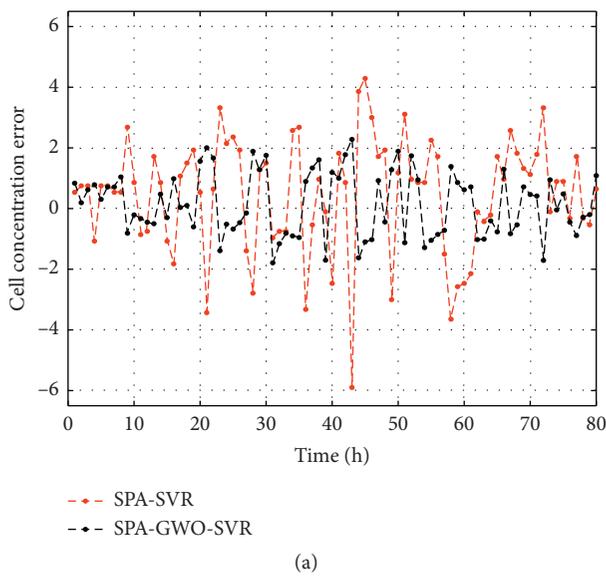


FIGURE 6: Continued.

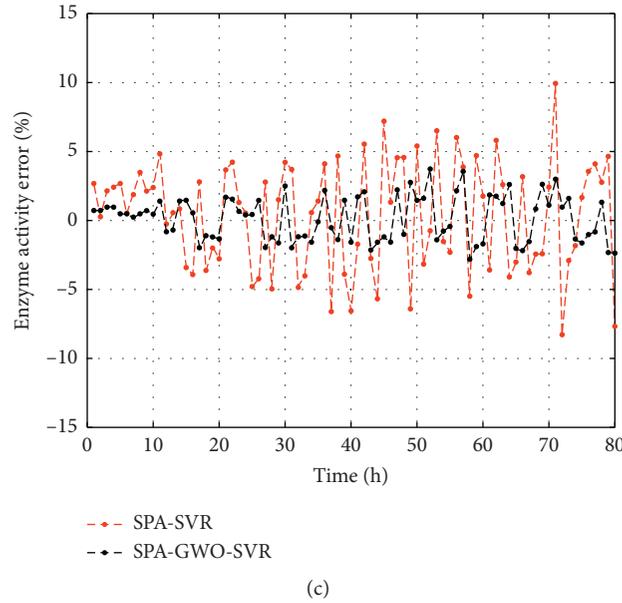


FIGURE 6: Comparisons of the predicted error curve. (a) Prediction error of cell concentration. (b) Prediction error of matrix concentration. (c) Prediction error of enzyme activity.

3.3. Simulation Analysis. To establish a multi-input and multi-output soft sensor model, by analyzing the results of Table 2, the p , DO , pH , v , μ , η , c_1 , c_2 variables are used as an input for the GRA-SVR model, and DO , pH , μ , η , c_1 are used as input variables for the SPA-SVR soft sensor model. The RMSE and coefficient determination R^2 are taken as the evaluation criteria, comparing the SVR model established by the full auxiliary variable and analyzing the model's performance based on GRA and SPA. It can be seen from Table 3 that, for each key variable, the prediction effect of the SPA variable selection method is significantly better than the model based on GRA and full auxiliary variables. It has a better variable selection effect, virtually eliminates redundant auxiliary variables, and reduces model complexity. But the prediction accuracy of the model still has room for improvement. Therefore, this paper employed the GWO algorithm to optimize the SVR model's critical parameters (σ , c , and ϵ). Set the grey wolf population size $N=50$, the maximum number of iterations $T=100$, and the optimal SVR parameter values $\sigma=81.31$, $c=45.52$, $\epsilon=5.13 \times 10^{-5}$, respectively. The optimized model results are shown in Table 4.

It can be observed clearly from Tables 3 and 4 that the proposed SPA-GWO-SVR soft sensor model provides a smaller RMSE value and a larger R^2 value than those of any individual SPA-SVR model which means the use of the above method can significantly improve the prediction results. The value of R^2 increased by 1.01%, 1.68%, and 1.43%, respectively; RMSE value decreased by 0.8847, 0.6132, and 0.0237, respectively. The optimization of parameters has achieved very significant results. Also, from the results of the training set and the test set, it can be found that as the number of predicted samples increases, the RMSE value also increases and R^2 decreases. Comparing the predicted result

between the SPA-GWO-SVR soft sensor model and simple SPA-SVR is shown in Figures 5 and 6. We comprehensively show that the soft sensor model based on SPA-GWO-SVR has better prediction performance.

4. Conclusion

Marine protease is a new type of enzyme preparation obtained from the fermentation of marine organisms. Because the ocean they live in has a particular environment, the enzymes they produce have more unique properties than those produced by other microorganisms, such as low-temperature resistance, alkali resistance, and pH . It has a wide range of action, making marine microorganisms have excellent development potential and application prospects in fermentation, such as food processing, enzyme industry, additives, and medicine. In marine microbial fermentation, to ensure the quality of fermentation products, it is necessary to detect a series of biological variables in real-time, especially the substrate concentration, cell concentration, and product concentration (enzyme activity). Currently, online measuring instruments can only detect certain physical and chemical variables in the fermentation process. There is no mature and practical instrument to measure these key biological variables. Therefore, in this paper, a soft sensor modeling method based on SPA-GWO-SVR was proposed. The auxiliary variables and target quality variables of the soft sensor model were determined based on SPA and GRA algorithm. The GWO algorithm is used to optimize the critical parameters (such as kernel function width σ , penalty factor c , and insensitivity coefficient ϵ) of the SVR model. Finally, the proposed SPA-GWO-SVR method is applied to predict the marine protease fermentation process's target quality variables. From the prediction results on the testing

dataset, the SPA-GWO-SVR soft sensor model's prediction performance is superior to GRA and SPA-SVR model.

This article has achieved specific results in the research of soft sensor modeling in fermentation. However, due to resource and capacity constraints, some shortcomings need to be further studied and perfected. In the future, it can be carried out from the following aspects. Although the soft sensor method proposed in this article is not limited to marine protease's fermentation process, it has an absolute versatility. The model parameters can also be modified and applied to other fermentation processes. Still, it needs further expansion and improvement to realize other biochemical processes' simulation training and enhance its universality. The sample data information obtained during the fermentation of the marine alkaline protease studied in the article is obtained based on a laboratory platform. It cannot completely replace the actual industrial production process.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was supported by the National Science Research Foundation of China (41376175), the Natural Science Foundation of Jiangsu Province (BK20140568, BK20151345), and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- [1] J.-H. Hao and M. Sun, "Purification and characterization of a cold alkaline protease from a psychrophilic *Pseudomonas aeruginosa* HY1215," *Applied Biochemistry and Biotechnology*, vol. 175, no. 2, pp. 715–722, 2015.
- [2] M. Saleem, A. Rehman, R. Yasmin, and B. Munir, "Biochemical analysis and investigation on the prospective applications of alkaline protease from a *Bacillus cereus* strain," *Molecular Biology Reports*, vol. 39, no. 6, pp. 6399–6408, 2012.
- [3] G. A. Montague, A. J. Morris, and J. R. Bush, "Considerations in control scheme development for fermentation process control," *IEEE Control Systems Magazine*, vol. 8, no. 2, pp. 44–48, 1988.
- [4] X. Zhu, K. Cai, B. Wang, and K. U. Rehman, "A dynamic soft sensor modeling method based on MW-ELWPLS in marine alkaline protease fermentation process," *Preparative Biochemistry & Biotechnology*, vol. 50, pp. 1–10, 2020.
- [5] X. Zhu, K. U. Rehman, B. Wang, and M. Shahzad, "Modern soft-sensing modeling methods for fermentation processes," *Sensors*, vol. 20, no. 6, p. 1771, 2020.
- [6] K. Pelckmans, J. De Brabanter, J. A. Suykens, and B. De Moor, "The differogram: non-parametric noise variance estimation and its use for model selection," *Neurocomputing*, vol. 69, no. 1–3, pp. 100–122, 2005.
- [7] W. Yan, H. Shao, and X. Wang, "Soft sensing modeling based on support vector machine and Bayesian model selection," *Computers & Chemical Engineering*, vol. 28, no. 8, pp. 1489–1498, 2004.
- [8] J. Grahovac, A. Jokić, J. Dodić, D. Vućurović, and S. Dodić, "Modelling and prediction of bioethanol production from intermediates and byproduct of sugar beet processing using neural networks," *Renewable Energy*, vol. 85, pp. 953–958, 2016.
- [9] X. Yuan, C. Ou, Y. Wang, C. Yang, and W. Gui, "A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2020.
- [10] Y. Liu, Z. Zhu, and X. Zhu, "Soft sensor modeling for key parameters of marine alkaline protease MP fermentation process," in *Proceedings of the 2018 Chinese Control and Decision Conference (CCDC)*, pp. 6149–6154, Shenyang, China, June 2018.
- [11] X. Yuan, S. Qi, Y. Wang, and H. Xia, "A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data," *Control Engineering Practice*, vol. 104, p. 104614, 2020.
- [12] X. Yuan, L. Li, Y. Shardt, Y. Wang, and C. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4404–4414, 2021.
- [13] X. Zhu, K. U. Rehman, W. Bo, M. Shahzad, and A. Hassan, "Data-driven soft sensor model based on deep learning for quality prediction of industrial processes," *SN Computer Science*, vol. 2, no. 1, pp. 1–10, 2021.
- [14] G. Wang, B. Xu, and W. Jiang, "SVM modeling for glutamic acid fermentation process," in *Proceedings of the 2016 Chinese Control and Decision Conference (CCDC)*, pp. 5551–5555, Yinchuan, China, May 2016.
- [15] B. Wang, M. Shahzad, X. Zhu, K. U. Rehman, M. Ashfaq, and M. Abubakar, "Soft-sensor modeling for l-lysine fermentation process based on hybrid ICS-MLSSVM," *Scientific Reports*, vol. 10, no. 1, pp. 1–15, 2020.
- [16] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 57, no. 2, pp. 65–73, 2001.
- [17] J. Zhang, B. Rivard, and D. Rogge, "The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data," *Sensors*, vol. 8, no. 2, pp. 1321–1342, 2008.
- [18] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 131–159, 2002.
- [19] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [20] B. Wang, M. Shahzad, X. Zhu, K. U. Rehman, and S. Uddin, "A non-linear model predictive control based on grey-wolf optimization using least-square support vector machine for product concentration control in L-lysine fermentation," *Sensors*, vol. 20, no. 11, p. 3335, 2020.
- [21] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Data preprocessing for supervised learning," *International Journal of Computer Science*, vol. 1, no. 4, pp. 30–37, 2006.
- [22] C. Zhu and D. Gao, "Influence of data preprocessing," *Journal of Computing Science and Engineering*, vol. 10, no. 2, pp. 51–57, 2016.