

Research Article

Optimization Modeling and Empirical Research on Gasoline Octane Loss Based on Data Analysis

Ji Guo ^{1,2} Yujia Lou ¹ Wanyi Wang ¹ and Xianhua Wu ^{1,2}

¹School of Economics and Management, Shanghai Maritime University, Shanghai 201306, China

²Collaborative Innovation Center on Climate and Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China

Correspondence should be addressed to Xianhua Wu; 185390@shmtu.edu.cn

Received 11 February 2021; Accepted 29 April 2021; Published 12 May 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Ji Guo 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.

Gasoline is one of the most consumed light petroleum products in transportation and other industries. This paper proposes a method for optimizing gasoline octane loss using data analysis technology aimed at optimizing the production process and minimizing the loss of gasoline octane. Firstly, the data are screened and the high-dimensional data are reduced to construct the neural network prediction model optimized by genetic algorithm. After utilizing the model for prediction, the optimal operating condition is achieved. Secondly, ensuring that the gasoline emission meets the standard, the octane loss is reduced by adjusting the operating variables. Thirdly, actual data are collected and calculated to obtain the main operating variables and their optimal operating conditions of a petrochemical company affecting the catalytic cracking gasoline S-Zorb unit, thus providing companies using S-Zorb units with reference data for optimizing gasoline catalytic cracking processes. Fourthly, the superiority of the proposed method was verified by comparing it with the other methods. This paper intends to contribute to better modeling the progress of gasoline catalytic cracking by adequately considering the impact of multiple factors, improving the quality of refined oil products of chemical enterprises, saving the economic cost of chemical enterprises, and protecting the atmospheric environment.

1. Introduction

As the world's most consumed light petroleum product, gasoline is one of the main fuels for automobiles. In 2018, global vehicle consumption is 9.5 million [1]. In the same year, the automobile industry consumes about 1.04 million ktoe of motor gasoline, while the transportation sector consumes 2.65 million ktoe oil products around the world, the automobile industry accounts for nearly half of the total consumption [2]. The exhaust gas emitted by gasoline combustion has a large negative impact on the atmospheric environment [3–6] and residents' health [7]. In 2018, greenhouse gas emissions from the transportation sector account for about 20% percent of the total global emissions [8]. Therefore, the cleaning of gasoline is an important task for all countries in the world to purify air pollution. An important issue for clean gasoline production is to maintain

the octane level of crude oil as its combustion performance index while reducing the sulfur and olefin content of crude oil so as to improve the quality of finished gasoline.

In the analysis and modeling of gasoline catalytic cracking, the high complexity of refining technology and the diversity of equipment have led to the nonlinear relationship between the operating variables of the control equipment and the strong coupling. The previous analysis models have fewer variables [9–13], and the models have higher requirements for raw materials [14–16], resulting in a lag in the response of the optimization process of the model; the refined oil obtained has more impurities and worse combustion performance, causing a certain economic loss. To avoid unnecessary losses, the modeling process needs to be optimized.

This article puts forward a novel idea: based on data analysis, the optimization process of gasoline catalytic

cracking was modeled and actual data were collected to verify the practicability of the idea. This article intends to contribute to better modeling progress of gasoline catalytic cracking by adequately considering the impact of multiple factors, to optimize the production process and minimize the loss of gasoline octane.

This paper has seven parts. Besides the introduction, the second part is a literature review; the third part introduces the research ideas, data sources, and data dimensionality reduction processing process and constructs the prediction model; the fourth part proposes the prediction results; the fifth part analyzes the economic benefits gained after reducing octane loss; the sixth part shows the robust test with the PCA method; and the seventh part gives out the conclusion and the direction of future research.

The research idea of this article is of the following steps:

- (1) *Data Collection.* Historical accumulated data of a petrochemical company's S-Zorb device are collected, including operating variable data, raw material data, product data, and catalyst data. After adjusting the data frequency difference, 325 samples are formed.
- (2) *Data Cleansing.* The maximum and minimum limit method, the PauTa criterion, and the null value processing method are used to deal with the values, with abnormal data and abnormal measured variables eliminated.
- (3) *Data Dimensionality Reduction.* The dimensionalities of the processed data are reduced by the locally linear embedding method. The measured variables are retained according to the displayed weights to ensure that the selected variables are representative. Then, these variables are tested for correlation coefficients to remove redundant variables with too high correlation coefficients to ensure the independence between variables.
- (4) *Model Building.* The aforementioned dimensionality reduction variables and octane loss value are employed, a BP neural network model based on the genetic algorithm is established, and the optimal octane loss prediction model is obtained after adjusting the parameters.
- (5) *Results Forecast.* Using the optimal octane loss prediction model, after forecasting and weighted average according to the desulfurization standard, the predicted optimal operating conditions of the main operating variables are obtained.
- (6) *Benefits Analysis.* With the predicted optimal operating conditions obtained, the cost saved by reducing octane loss is compared with the cost of original operating conditions to clearly show the economic benefits.
- (7) *Robust Test.* The PCA method is employed in the step of data dimensionality reduction.

Then, the GA-BP model is also constructed, and the fitting degree is tested for the robust test.

The process is shown in Figure 1.

This article puts forward an application-oriented method. Compared with previous studies, the possible innovations in this article are as follows:

- (1) A locally linear embedding dimensionality reduction algorithm is used to reduce the dimensionalities of high-dimensional data. Traditional linear dimensionality reduction methods, such as principal component analysis and factor analysis, are not appropriately suitable for these high-dimensional data. The locally linear embedding used in this paper can map high-dimensional variable data to low-dimensional vector space based on maintaining the relationships among data, to reduce the dimensionalities and optimize the modeling process.
- (2) A BP neural network model optimized based on a genetic algorithm is used to predict gasoline octane loss. Previous studies are mostly implemented from the perspective of chemical mechanism, reducing the octane loss of refined oil by adjusting device configuration or chemical process, raw material properties, and so on. On the other hand, for a large number of variables and time-series data provided by the factory, the traditional regression analysis is hard to be applied because the degree of freedom loss is serious. This paper analyzes the impact of operating variables on gasoline octane loss through data analysis technology and provides enterprises with operating conditions to reduce gasoline octane loss from a new perspective and can further optimize the operation plan based on the mechanism of refined oil production.

2. Literature Review

The cleaning of finished products of petroleum resources and the maximization of efficiency are currently problems facing the chemical industry. In terms of the cleaning of finished products, the main indicator is to reduce the sulfur and olefin content in the finished products. In terms of efficiency maximization, the main indicator is the retention of the octane number (Research Octane Number, RON) in the finished product, which is the most important indicator reflecting the combustion performance of gasoline and is also the brand name of gasoline.

The catalytic cracking technology used for refining petroleum is to crack heavy oil through the combined action of heat energy and catalyst to transform it into cracked gas, gasoline, and diesel [17]. During the catalytic cracking process, flue gas desulfurization is used [18–20], with olefins converted into liquefied petroleum gas by-products. The light oil using this technology has high productivity and good octane retention. Further improvement of catalytic cracking technology is conducive to better cleaning of finished products and maximum efficiency.

In the past, research on catalytic cracking technology focused mainly on mechanism models and technical improvements. Salazar et al. [21] proposed a process for upgrading nitrogen-rich and sulfur-rich heavy oil feedstock, which reduces the nitrogen and sulfur content while increasing the octane number. Valla et al. [15] studied the influence of various types of catalytic cracking feedstock on the distribution of

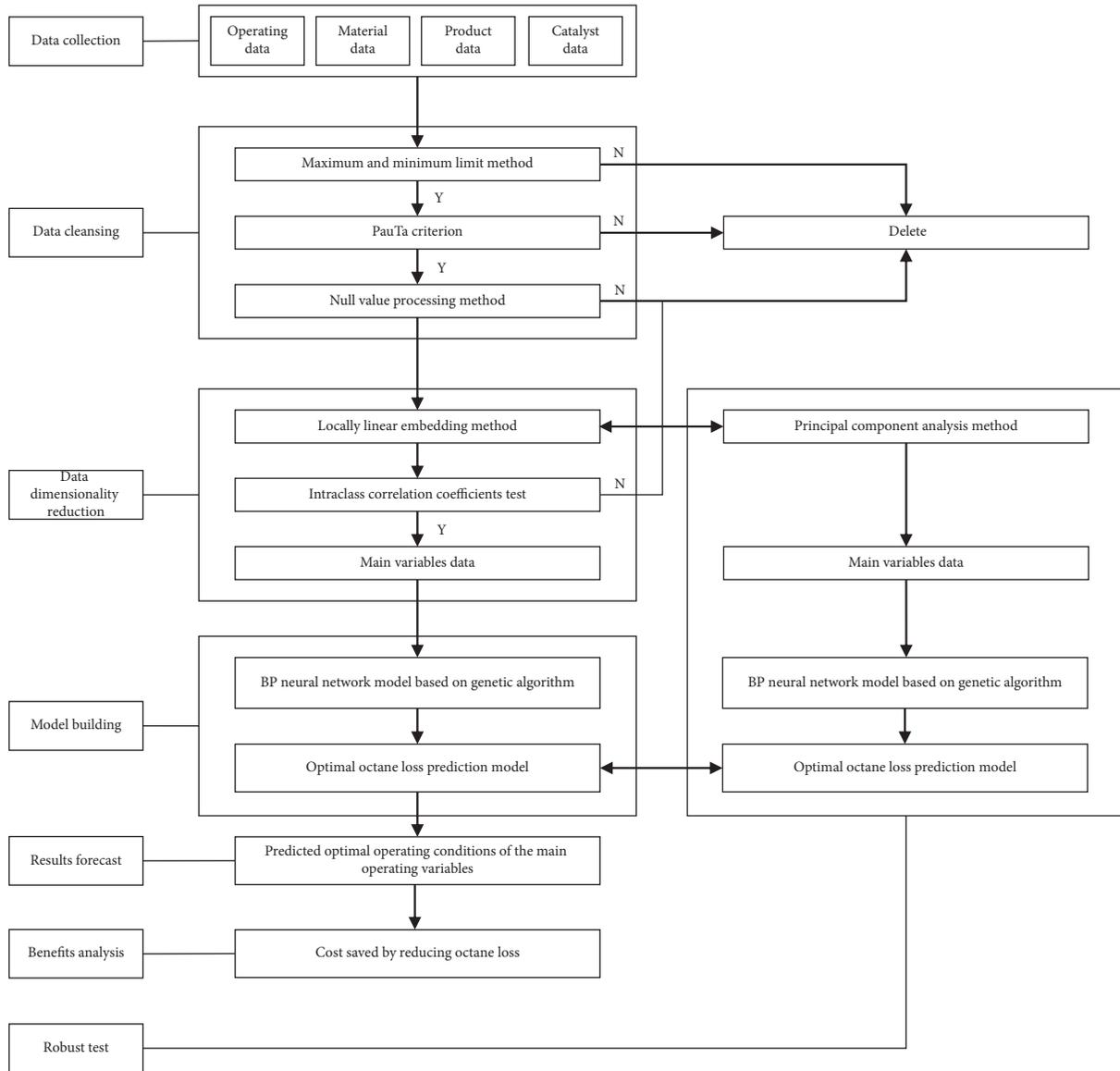


FIGURE 1: The flow chart of optimization modeling on gasoline octane loss.

sulfides in finished products and proposed the formation mechanism of sulfides on this basis. Brunet et al. [14] investigated the possible sources of sulfur impurities, discussed various factors affecting the hydrodesulfurization and olefin hydrogenation reactions, such as catalysts, carrier properties, and additives, and introduced processes for preserving the octane number of FCC gasoline. Li et al. [22] proposed selective hydrodesulfurization (RSDS-I) technology, which showed the superior desulfurization ability in industrial applications. Li et al. [23] chose $\text{CoMoP}/\eta\text{-Al}_2\text{O}_3$ as catalyst to reduce sulfur content of coal tar light oil (CTLO), which is a potential material for the manufacture of high-octane gasoline blending component. Ayoub and Masoud [24] carried out the development of hydrodesulfurization as an alternative for cleaner production of liquefied petroleum gases. Yang et al. [16] analyzed the main factors affecting the change of gasoline octane number, such as the nature of raw materials, catalysts, and device operating conditions. Hasheminejad et al. [25] designed

a new material for adsorptive desulfurization to achieve the lowering sulfur level in fuels. Qin et al. [26] established a model for the FCC process on a molecular level employing the structure-oriented lumping (SOL) method to investigate the effects of the diameter expanding reactor. The model established could calculate the molecular level product distribution from the reactor inlet to the outlet to reduce the olefins content and improve the iso-paraffins content of gasoline. Li et al. [27] put forward a high-efficient optimization of corresponding operation conditions for olefin separation. They researched solvent extraction separating olefin and extracted sulfides simultaneously to protect the RON from a loss during hydrodesulfurization.

There are also scholars conducting research on the use of data analysis techniques to establish related models. Qin and Chen [11] proposed a neural network prediction model for gasoline octane number, but only three influencing variables were considered: temperature, pressure, and flow rate of

continuous reforming reactor. Wei [28] used the principal component analysis and BP (back propagation) neural network to process the gas sensor array signal and analyzed the response of the array to gasoline, ethanol, and their mixtures. Yang et al. [13] used a neural network model optimized by the genetic algorithm BP (GA-BP) to establish a gasoline blending model. Paranghooshi et al. [10] used the artificial neural network (ANN) model to determine the octane number of gasoline blends produced by Tabriz refinery. Cheng and Yi [29] proposed a neural network model-based predictive control method for the etherification of FCC light oil. Zhang et al. [30] compared the accuracy of the BP model and the GA-ANN model in predicting the gasoline output of the RFCCU unit. Su et al. [31] used the GA-BP model to predict the production of coke as the main by-product in the catalytic cracking reaction. Ouyang et al. [9] used 19 input variables to establish a BP neural network and studied the influence of raw material preheating temperature, two reaction zone outlet temperatures, and reaction pressure on product distribution. They also used genetic algorithms to optimize the operating variables, and the gasoline yield was significantly improved after optimization. Tian et al. [12] used the particle swarm optimization (PSO-BP) neural network to predict the law of sulfur content in refined diesel products with operating parameters and compared the performance of BP, GA-BP, and PSO-BP. Cheng et al. [32] combined the BP neural network with the PID control, optimized the parameters of the controller, and applied it to the flow control of catalytic cracking natural gas, so as to control the concentration of the final product of heavy oil. Zhang et al. [33] established a nonrandom two-liquid model and a simulation method for the FCC naphtha solvent extraction process to fully improve the structure of the components of FCC naphtha aimed at the production of ultra-low-sulfur gasoline with minimal loss of octane number. Foroughi et al. [34] applied an artificial neural network to predict the volume percentage of adulterants to protect adulteration in gasoline based on the set of an automatic distillation apparatus measuring the recovered volume and temperature. Ma et al. [35] clarified the mechanism of research octane number (RON) and motor octane number (MON) of gasoline influenced by initial thermodynamic conditions and fuel chemistry. They proposed a hybrid analysis framework, with a combination of the transient tracking method and data-driven modeling algorithm to realize the fast-forward prediction and analysis on fuel's ONs.

In summary, in terms of optimizing the process of gasoline catalytic cracking, the academic circles mainly modify the raw material properties and optimize the steps and the equipment in the gasoline catalytic cracking process from the perspective of chemical industry. When using data analysis to adjust the corresponding catalytic cracking operation variables to achieve optimization, most studies selected few variables, and it is difficult to fully consider the impact of multiple factors. This article hopes to contribute in this regard.

3. Data and Methodology

3.1. Data Sources. The original data are the historical accumulated data of the catalytic cracking gasoline S-Zorb unit

of a petrochemical company. Operating variable data come from real-time database. The collection time is from April 2017 to May 2020, with a total of 353 operating variable sites collected. From April 2017 to September 2019, the data collection frequency was 3 minutes/time; from October 2019 to May 2020, the data collection frequency was 6 minutes/time. The data time range for raw materials, products, and catalysts was from April 2017 to May 2020. The octane number of raw materials and products is an important modeling variable. Since the octane number is difficult to measure, the data collection frequency is twice a week. The difference in the frequency of data collection is adjusted according to the actual situation. The measured value of the octane number can be regarded as the comprehensive effect within two hours before the moment. The value of the manipulated variable is the average value of the previous two hours, which corresponds to the measured value of the octane number at that moment, resulting in 325 samples with a fixed time interval of two weeks (see Supplemental materials: S-Table 1).

3.2. Data Cleansing. Since the enterprise device has been operating continuously for 4 years, it is necessary to improve the accuracy and effectiveness of the recorded data. Before modeling, the data need to be sorted out. In the original data of the sample, most of the variable data is normal, but some data of each device have problems, some variables only contain data for part of the period, and some of the variables' data are all null values or some of the data are null values. Therefore, this article will process the original data for subsequent research.

First, according to the chemical process requirements and operating experience, the original data variables have a certain operating range (see Supplemental materials: S-Table 2), with the maximum and minimum limiting method used to remove some samples that are not within this range.

Second, outliers are removed according to the PauTa criterion (the 3δ criteria). The variable is measured with equal precision, with x_1, x_2, \dots, x_n obtained, where n is the number of samples. The arithmetic mean \bar{x} and the residual error $v_i = x_i - \bar{x}$ ($i = 1, 2, \dots, n$) are calculated, with the standard error δ determined according to Bessel's formula as follows:

$$\delta = \left(\frac{\sum_{i=1}^n v_i^2}{n-1} \right)^{(1/2)} = \left\{ \frac{1}{n-1} \cdot \left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right] \right\}^{(1/2)}. \quad (1)$$

If the residual error v_b ($1 \leq b \leq n$) of a certain measured value x_b satisfies $|v_b| = |x_b - \bar{x}| > 3\delta$, x_b is considered to be a bad value with a larger error value and is eliminated.

Finally, the null values in the measured variables of the data sample are processed. After counting, the average of the number of null values in each column is calculated and defined as the critical point for judging incomplete data. If the number of null values in the measured variable exceeds this critical value, it is considered that there are too many

incomplete data in this column of data, which will adversely affect the goodness of fit, so this type of measured variable is eliminated. After calculation, the critical point of incomplete data judgment is 104, and 16 measured variables are eliminated after processing (see Table 1).

3.3. Data Dimensionality Reduction. When modeling, the variables need to be reduced in dimensionality. Dimensionality reduction is conducive to screening out the operating variables having the greatest impact on the octane number in the catalytic cracking process, ignoring secondary variables, and improving application efficiency. Due to the large number of variables, they can be regarded as high-dimensional data; there is a nonlinear relationship between each other; the nonlinear dimensionality reduction algorithm is more suitable for the above situation.

3.3.1. LLE Algorithm. This paper uses locally linear embedding (LLE) [36, 37] in the nonlinear dimensionality reduction algorithm for data dimensionality reduction. Compared with the traditional PCA (principal component analysis) sample variance reduction method, it retains the local linear characteristics of the sample when reducing the dimensionality. The principle of LLE is that high-dimensional data are approximately locally linear in a very small local neighborhood in Euclidean space, and that a certain point can be represented by linear least squares of its surrounding points. LLE uses the linear fitting coefficients as the local geometric properties of the point to find the low-dimensional projection of the data. This algorithm is widely used in the field of high-dimensional data (see Appendix A for the specific algorithm).

3.3.2. Data Dimensionality Reduction Results. After obtaining the weight coefficient matrix through the LLE algorithm, the 29 control variables with the highest weights in the weight coefficient matrix are extracted. Taking into account the actual situation in the chemical process, the "Octane number of raw materials" should be taken as one of the main variables and included in the analysis, resulting in 30 main variables.

The use of the LLE algorithm for dimensionality reduction can ensure that the above main variables are representative and then can determine whether the variables are related by calculating their correlation coefficients. The thermodynamic diagram representing the correlation coefficient of the variable is shown in Figure 2.

It can be seen from Figure 2 that the correlation between some variables and other variables is too high (the absolute value of the correlation coefficient being greater than 0.8), so the seventh, 12th, and 18th variables are deleted. After the deletion, the correlation coefficient of the representative variables decreases (as shown in Figure 3), which ensures the independence between the representative variables.

After processing and testing, 27 variables were finally retained as main variables, with the specific main variables shown in Table 2.

3.4. GA-BP Model Establishment. With the development of intelligent machine algorithms, data analysis techniques can be used to solve the problem of excessively high data dimensions. This paper establishes a GA-BP neural network model to predict the loss of gasoline octane number, aiming to adjust operating variables and reduce the loss of octane number in the chemical process.

3.4.1. Model Design. The back propagation (BP) neural network [38–45] is based on intelligent machine learning, has nonlinear features, good classification ability, and mapping ability to multidimensional functions, and has great advantages in multivariate regression. It has an input layer, an intermediate layer, and an output layer. In essence, the square of the network error is used as the objective function, and the gradient descent method is used to obtain the minimum value of the objective function. The calculation process is divided into two parts: forward propagation and backward propagation. When calculating the error in the forward direction, it goes from the input layer to the output layer, and the reverse adjustment process is to adjust the weight and threshold of the network through the distribution of the error signal in each layer so that the network error decreases along the gradient direction. (Figure 4)

The genetic algorithm (GA) [42, 43] is a parallel random search optimization method based on the theory of biological evolution to simulate the genetic mechanism of nature. Based on the principle of "survival of the fittest" in nature, individuals are selected, crossed, and mutated; individuals are screened according to the selected fitness function, individuals with better fitness are retained, and individuals with poor fitness are eliminated. The new group is better than the previous generation on the basis of inheriting the information of the previous generation. They loop iteratively until the optimization is reached.

As a local search optimization method, the BP neural network is easy to fail in the case of nonlinear complex problems. In addition, the BP algorithm is also prone to overfitting. The genetic algorithm can optimize the neural network learning rules and improve the computational efficiency of the neural network; it is necessary to build a dynamic neural network structure to avoid the final result of the model being a local optimum instead of a global optimum. Therefore, a BP neural network model is employed optimized based on a genetic algorithm, namely, the GA-BP model [44] to predict product octane loss (see Appendix B for specific algorithm design).

3.4.2. Model Results. The input layer X of the neural network is the 27 operating variables screened out above, the output layer P is the octane loss value and the product sulfur content 2 variables, and the formula of the hidden layer node is $Y = \sqrt{n_x + n_p} + L$, ($1 \leq L \leq 10$), which is set to 7 here. In order to improve the accuracy of neural network prediction, the number of the training set is set to 300 and the number of the test set is set to 25. Each iteration is randomly sampled from the sample set.

TABLE 1: The eliminated variables.

Order	Variable code	Variable name
1	S-ZORB.FC_2301.PV	D105 fluidized hydrogen flow rate
2	S-ZORB.FT_1501.PV	New hydrogen intake flow rate
3	S-ZORB.FT_5104.PV	Light hydrocarbon discharge flow rate
4	S-ZORB.FT_9101.PV	Oil discharge flow rate
5	S-ZORB.FT_1002.PV	# catalytic gasoline intake flow rate
6	S-ZORB.FC_1202.PV	D121 top remove flare flow rate
7	S-ZORB.FC_3103.PV	Regenerated cold nitrogen flow rate
8	S-ZORB.FT_1002.TOTAL	S-ZORB.FT_1002cumulative flow rate
9	S-ZORB.FT_1501.TOTAL	New hydrogen intake accumulated flow
10	S-ZORB.FT_5102.PV	S-ZORB.FT_5102 intake flow rate
11	S-ZORB.FT_2901.DACA	D-109 loose wind flow rate
12	S-ZORB.FC_1104.DACA	Feed control valve bypass flow rate
13	S-ZORB.FT_2803.DACA	D-102 emergency hydrogen flow rate
14	S-ZORB.FT_1502.DACA	Supply hydrogen compressor outlet return pipe flow rate
15	S-ZORB.TEX_3103A.DACA	Eh-102 heating element/a beam temperature
16	S-ZORB.FT_5102.DACA. PV	D-201 sulfur-containing effluent discharge flow rate

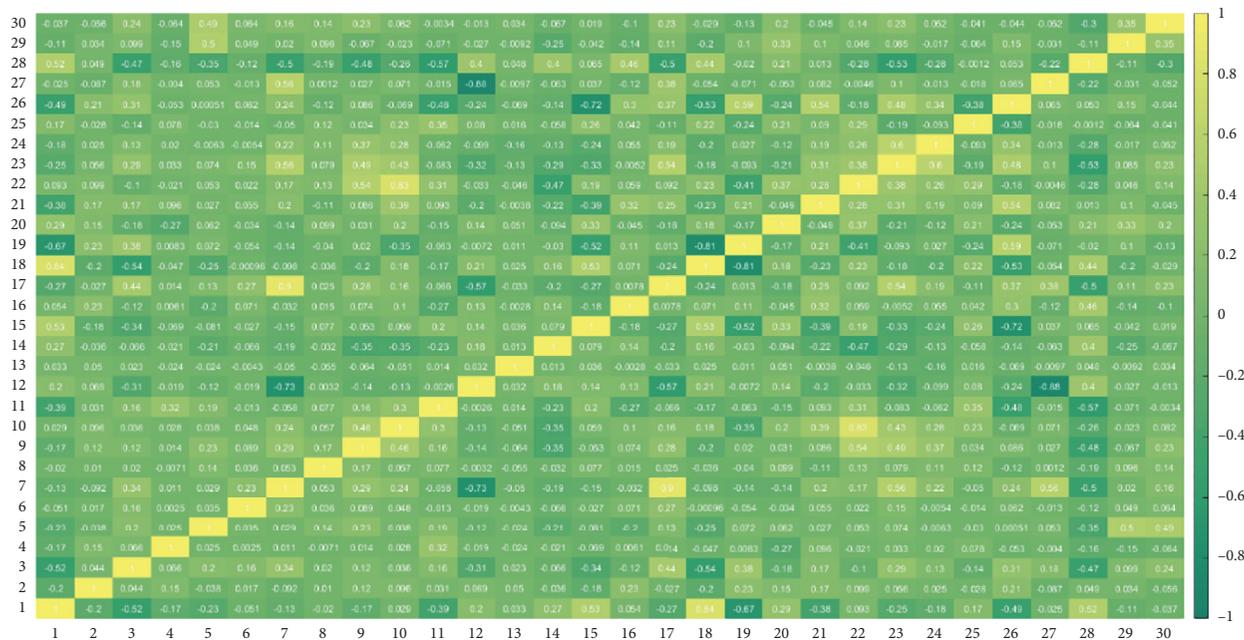


FIGURE 2: Main variables correlation coefficient thermodynamic diagram before deletion.

After selecting and adjusting the parameters, the genetic algorithm finally sets the parameters as follows: the maximum number of iterations G_{max} is set to 25, the population size M to 60, the crossover probability P_c to 0.2, and the mutation probability P_m to 0.05. The final parameters of the neural network are set as follows: the learning rate h is 0.1, the minimum error threshold E_{min} is set to 0.00001, and the maximum number of iterations G_{max} is set to 100.

After model construction, the overall goodness of fit of GA-BP prediction is 52.51%, the prediction error of octane number loss is small, and the prediction error of product sulfur content is large. The error, error percentage, and fitness curve between the predicted value and the true value are shown in Figures 5–7. The optimized neural network model after training is saved, which is the prediction model of gasoline octane number loss.

It is assumed that the physical and chemical properties of raw materials and standby and regenerated adsorbents in the optimization process remain unchanged (the octane number of raw materials and the sulfur content of raw materials in the main variables also remain unchanged). According to China's current standard GB 18352.6-2016, the sulfur content of finished gasoline products is required to be not more than $10 \mu\text{g/g}$. In order to leave room for business operations, the selection criteria for sulfur content in this article are not more than $5 \mu\text{g/g}$. For the 25 main operable variables corresponding to the above samples, simulation samples are generated according to the variable range and minimum change value, simulated the debugging process of the variables, and used the octane loss reduction as the weighted average to obtain the optimal operating conditions.

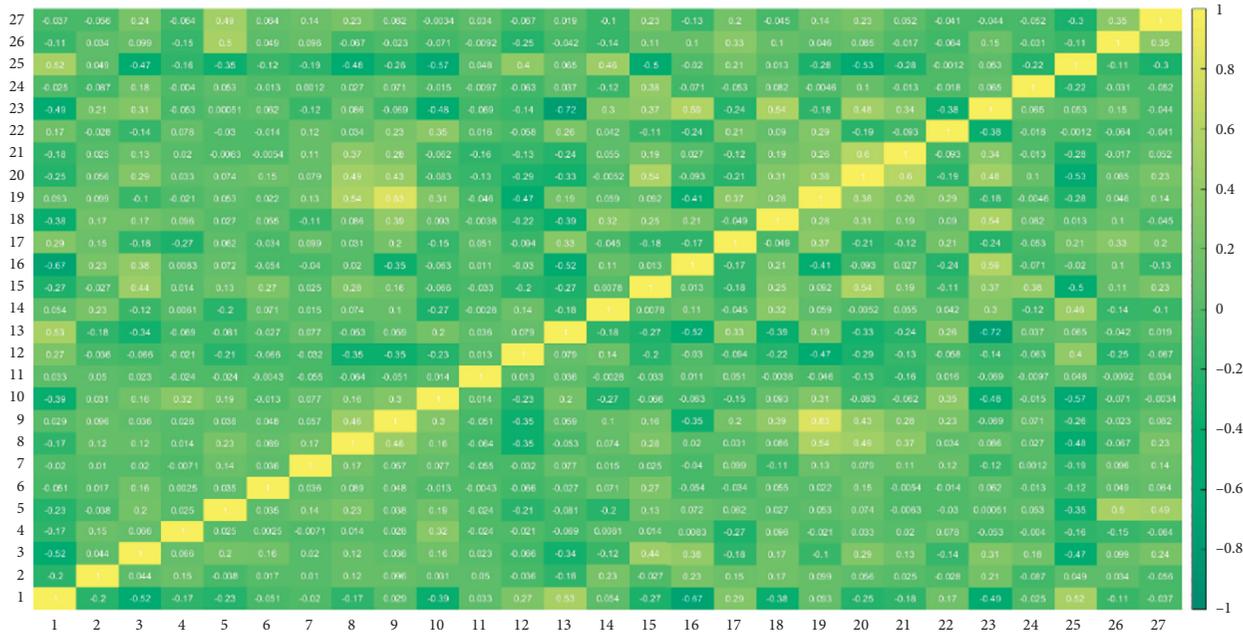


FIGURE 3: Main variable correlation coefficient thermodynamic diagram after deletion.

TABLE 2: Main variables after deletion.

Order	Variable code	Variable name
1	S-ZORB.TE_2301.PV	D105 temperature
2	S-ZORB.TE_9301.PV	1.0 MPa steam inlet temperature
3	S-ZORB.FT_3301.PV	Deaerated water inlet flow rate
4	S-ZORB.FT_9401.PV	Purify air inlet flow rate
5	S-ZORB.FT_1004.PV	3# catalytic gasoline intake flow rate
6	S-ZORB.AI_2903.PV	Oxygen content of regenerated flue gas
7	S-ZORB.LI_9102.DACA	D204 liquid level
8	S-ZORB.TE_9002.DACA	D203 top outlet pipe temperature
9	S-ZORB.TE_1502.DACA	D122 outlet pipe temperature
10	S-ZORB.FT_2302.DACA	D105 upper jumper loose wind flow rate
11	S-ZORB.SIS_PT_2602.PV	Regenerator top/regenerator receiver differential pressure
12	S-ZORB.PDT_3602.DACA	Cold nitrogen filter ME-114 differential pressure
13	S-ZORB.FT_3701.DACA	Closed hopper N2 filter outlet gas flow rate
14	S-ZORB.FT_3702.DACA	Lock hopper H2 filter outlet gas flow rate
15	S-ZORB.PDT_2605.DACA	R-102 bottom nozzle differential pressure
16	S-ZORB.PDT_2906.DACA	Me-108 filter differential pressure
17	S-ZORB.TE_7106.DACA	K-101a left exhaust temperature
18	S-ZORB.HIC_2533.AUTOMANA.OP	HV2533 hand operator
19	S-ZORB.TE_5009.DACA	E-205 inlet pipe temperature
20	S-ZORB.AT-0003.DACA.PV	S-ZORB.AT-0003
21	S-ZORB.AT-0010.DACA.PV	S-ZORB.AT-0010
22	S-ZORB.FT_5204.DACA.PV	Degassing of gasoline products
23	S-ZORB.FT_1006.TOTALIZERA.PV	Hydrocracking light naphtha intake accumulated flow
24	S-ZORB.FT_1503.DACA.PV	8.0 MPa hydrogen to circulating hydrogen compressor intake flow rate
25	S-ZORB.FT_1504.TOTALIZERA.PV	8.0 MPa hydrogen to back blow hydrogen compressor outlet accumulated flow
26	RON	Octane number of raw materials
27	S-ZORB.AT_1001.DACA	Sulfur content of raw gasoline

4. Main Results

In summary, we have obtained the main operating variables of the petrochemical enterprise that affect the catalytic cracking gasoline S-Zorb unit, as well as the optimal

operating conditions when the product sulfur content is less than 5 $\mu\text{g/g}$, as shown in Table 3.

The above results provide companies using S-Zorb with reference data for optimizing the process of catalytic cracking gasoline.

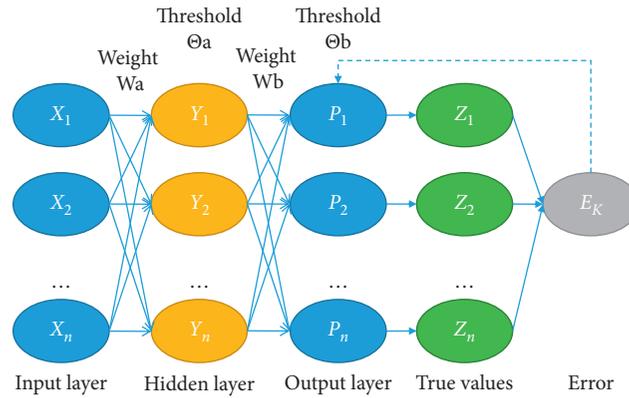


FIGURE 4: BP neural network structure.

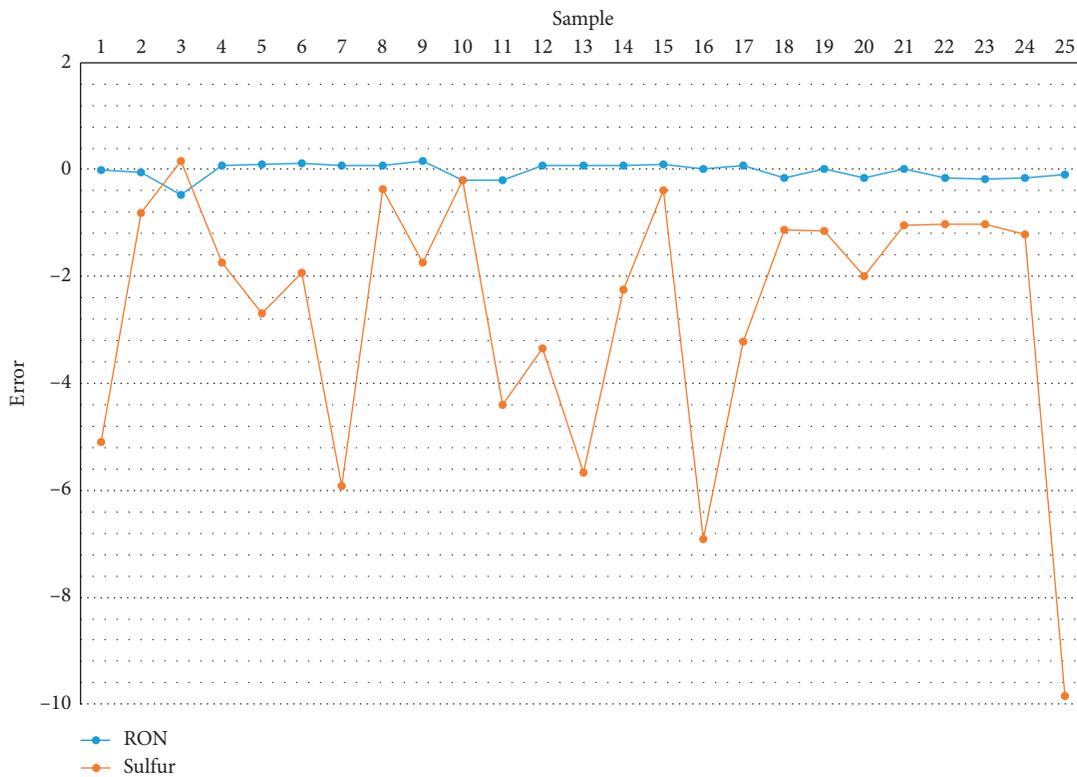


FIGURE 5: Error between the predicted value and the true value.

5. Economic Benefit Analysis

In the process of refining gasoline with S-Zorb unit in existing petrochemical enterprises, if the octane number is reduced by 1 unit, it is equivalent to reducing the economic loss of \$23.055 per ton.

Therefore, in this study, the average RON of the original data sample is 88.45 units, and the average octane number of the product under the optimal operating conditions is 89.70 units. After the prediction of the GA-BP model and the optimization of operating conditions, the octane number loss of the product is reduced by 1.25 units. If we take the automobile industry in 2018 as an example and calculate by

the annual gasoline consumption of 1.04 million kt, the application of this research can save about 29.97 million dollars for automobile industry in the whole world.

It can be seen that reducing the octane loss of gasoline products will bring huge economic cost savings to enterprises and will also cut down air pollution and the corresponding cost of treatment, so the conclusions of this study can provide excellent economic benefits in practice.

6. Robust Test

To verify the reliability of the method and conclusion in this paper, the principal component analysis (PCA) [46] method

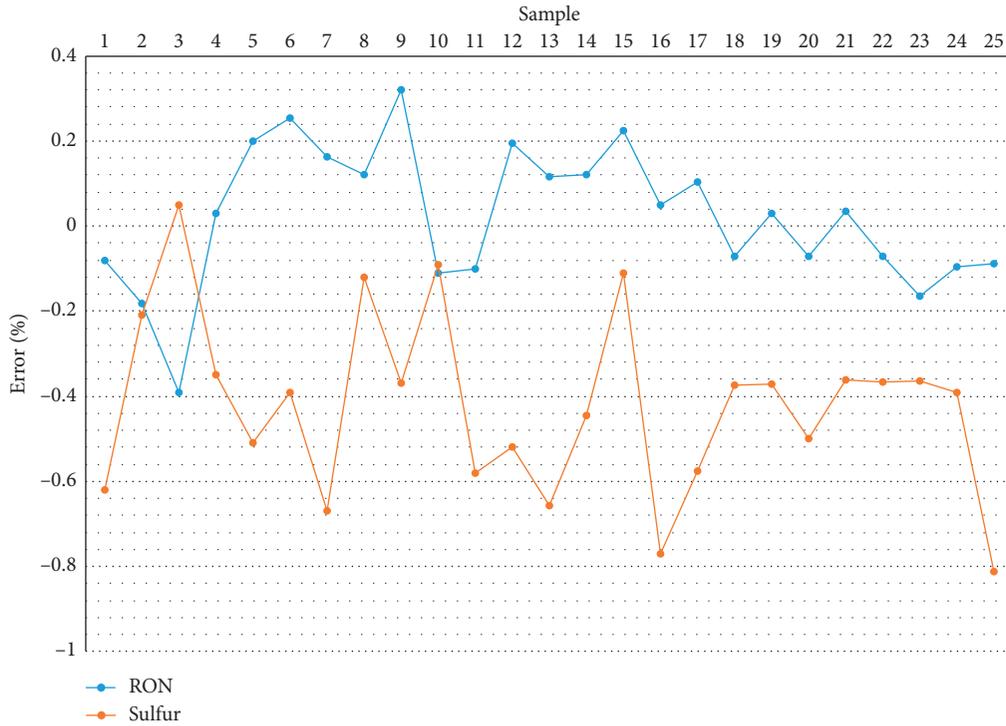


FIGURE 6: Error percentage between the predicted value and the true value.

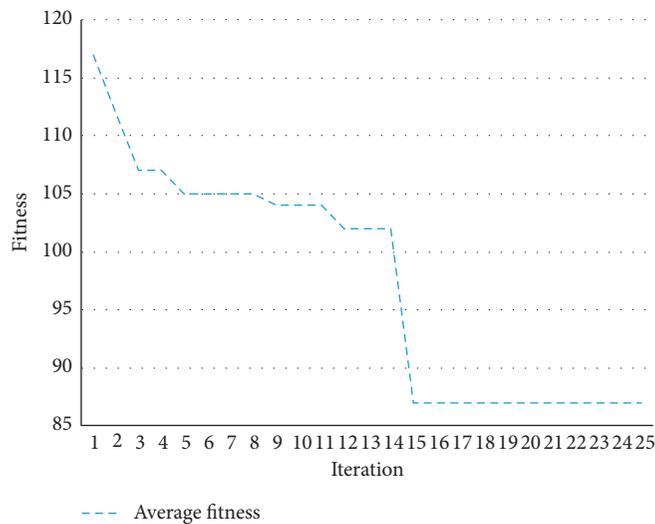


FIGURE 7: The fitness curve.

is used to reduce the dimension of the original variables and sample data, and then the GA-BP model is also constructed, and its fitting degree is tested for the robust test.

PCA aims at trying to create a new set of unrelated comprehensive indicators composed of the original indicators with different weights. Since it is necessary to adjust the operating variables of the S-Zorb device to achieve the goal of reducing octane number loss, the first 26 variables are also retained according to the weight and the variable “Octane number of raw materials” is added. For the variables retained that are already representative and unrelated, the

calculation of their correlation coefficients is omitted. The specific main variables are shown in Table 4.

Then, the main variables above are employed by GA-BP modeling. After model construction, the overall goodness of fit is 45.36% with a similar error performance. The error, error percentage, and fitness curve between the predicted value and the true value are shown in Figures 8–10 .

It can be concluded that the overall fit goodness of the PCA method is inferior to that of the LLE method. For this study focused on more than 300 variables, the PCA method is not appropriately suitable for such high-dimensional data.

TABLE 3: The optimal operating conditions of the main operating variables.

Order	Variable code	Variable name	Optimal operating conditions	Unit
1	S-ZORB.TE_2301.PV	D105 temperature	277.5374	°C
2	S-ZORB.TE_9301.PV	1.0 MPa steam inlet temperature	194.7160	°C
3	S-ZORB.FT_3301.PV	Deaerated water inlet flow rate	589.8554	kg/h
4	S-ZORB.FT_9401.PV	Purify air inlet flow rate	392.3236	Nm ³ / h
5	S-ZORB.FT_1004.PV	3# catalytic gasoline intake flow rate	37.6089	T/h
6	S-ZORB.AI_2903.PV	Oxygen content of regenerated flue gas	0.9389	%
7	S-ZORB.LI_9102.DACA	D204 liquid level	45.7631	%
8	S-ZORB.TE_9002.DACA	D203 top outlet pipe temperature	29.7332	°C
9	S-ZORB.TE_1502.DACA	D122 outlet pipe temperature	44.6834	°C
10	S-ZORB.FT_2302.DACA	D105 upper jumper loose wind flow rate	102.2920	Nm ³ / h
11	S-ZORB.SIS_PT_2602.PV	Regenerator top/regenerator receiver differential pressure	-126.1370	kPa
12	S-ZORB.PDT_3602.DACA	Cold nitrogen filter ME-114 differential pressure	0.4967	kPa
13	S-ZORB.FT_3701.DACA	Closed hopper N2 filter outlet gas flow rate	25.7114	Nm ³ / h
14	S-ZORB.FT_3702.DACA	Lock hopper H2 filter outlet gas flow rate	25.4971	Nm ³ / h
15	S-ZORB.PDT_2605.DACA	R-102 bottom nozzle differential pressure	12.1455	kPa
16	S-ZORB.PDT_2906.DACA	Me-108 filter differential pressure	10.3282	kPa
17	S-ZORB.TE_7106.DACA	K-101a left exhaust temperature	32.1123	°C
18	S-ZORB.HIC_2533.AUTOMANA.OP	HV2533 hand operator	76.8424	°C
19	S-ZORB.TE_5009.DACA	E-205 inlet pipe temperature	27.0397	°C
20	S-ZORB.AT-0003.DACA.PV	S-ZORB.AT-0003	3.0541	—
21	S-ZORB.AT-0010.DACA.PV	S-ZORB.AT-0010	1.0862	—
22	S-ZORB.FT_5204.DACA.PV	Degassing of gasoline products	1075.6240	kg/h
23	S-ZORB.FT_1006.TOTALIZERA.PV	Hydrocracking light naphtha intake accumulated flow	63021560.0446	kg/h
24	S-ZORB.FT_1503.DACA.PV	8.0 MPa hydrogen to circulating hydrogen compressor intake flow rate	2265244.5657	Nm ³ / h
25	S-ZORB.FT_1504.TOTALIZERA.PV	8.0 MPa hydrogen to back blow hydrogen compressor outlet accumulated flow	17278642.1169	Nm ³ / h

TABLE 4: Main variables retained by PCA.

Order	Variable code	Variable name
1	S-ZORB.TE_2002.DACA	R-101 bed middle temperature
2	S-ZORB.FT_1002.PV	1# catalytic gasoline intake flow rate
3	S-ZORB.FT_2901.DACA	D-109 loose wind flow rate
4	S-ZORB.FT_1002.TOTAL	S-ZORB.FT_1002 cumulative flow rate
5	S-ZORB.FT_3201.DACA	D-110 top pressure
6	S-ZORB.AT-0008.DACA.PV	S-ZORB.AT-0008
7	S-ZORB.PT_2106.DACA.PV	Back blow pressure
8	S-ZORB.PC_3501.DACA	Emergency hydrogen main
9	S-ZORB.AT-0011.DACA.PV	S-ZORB.AT-0011
10	S-ZORB.AT-0009.DACA.PV	S-ZORB.AT-0009
11	S-ZORB.AT-0007.DACA.PV	S-ZORB.AT-0007
12	S-ZORB.FT_9402.TOTAL	S-ZORB.FT_9402.TOTAL accumulated flow
13	S-ZORB.TXE_2203A.DACA	Eh-101 heating element temperature
14	S-ZORB.LC_5101.PV	Top reflux tank D201 level
15	S-ZORB.FT_5204.TOTALIZERA.PV	Gasoline product degassing accumulated flow
16	S-ZORB.TE_7508.DACA	K-103a exhaust temperature
17	S-ZORB.PT_7103.DACA	K-101a intake pressure
18	S-ZORB.PT_7107.DACA	K-101a exhaust pressure
19	S-ZORB.TC_2201.PV	Eh-101 inlet flow
20	S-ZORB.PT_7510.DACA	K-103a exhaust pressure
21	S-ZORB.SIS_PT_2703	D-110 bottom pressure
22	S-ZORB.FT_1202.TOTAL	S-ZORB.FT_1202.TOTAL accumulated flow
23	S-ZORB.SIS_LT_1001.PV	Material buffer tank liquid level
24	S-ZORB.TE_7102.DACA	K-101a intake temperature
25	S-ZORB.PT_7508.DACA	K-103a inlet pressure
26	S-ZORB.TE_7506.DACA	K-103a inlet temperature
27	RON	Octane number of raw materials

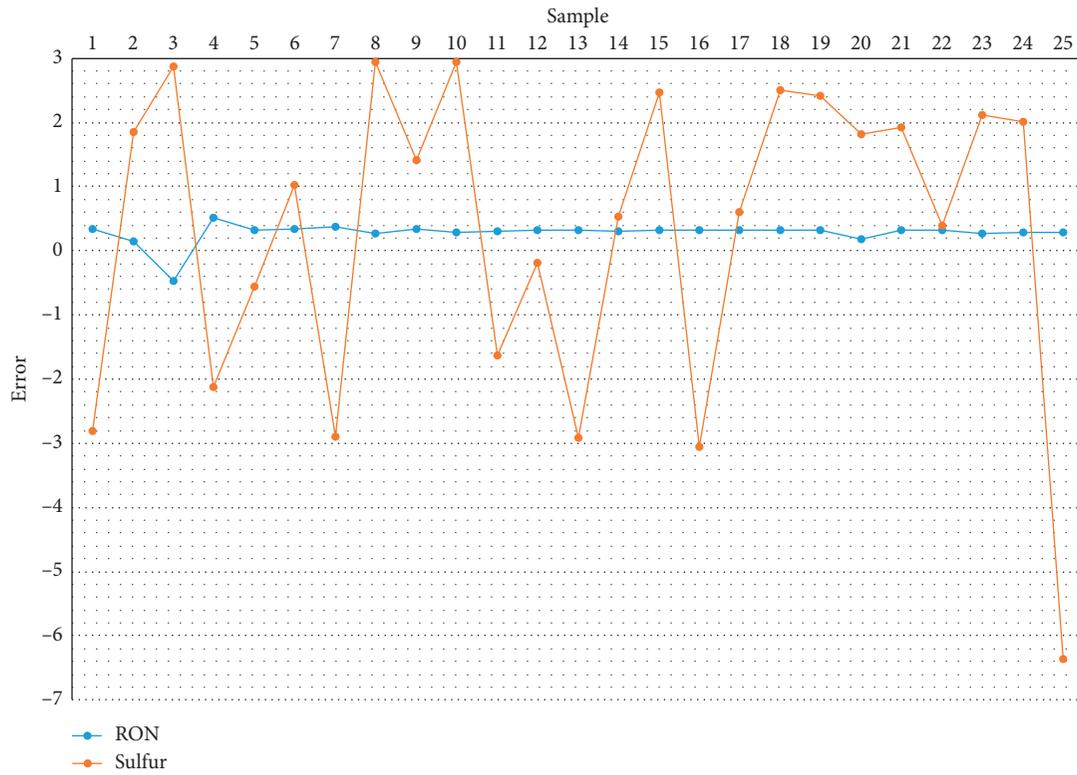


FIGURE 8: Error between the predicted value and the true value.

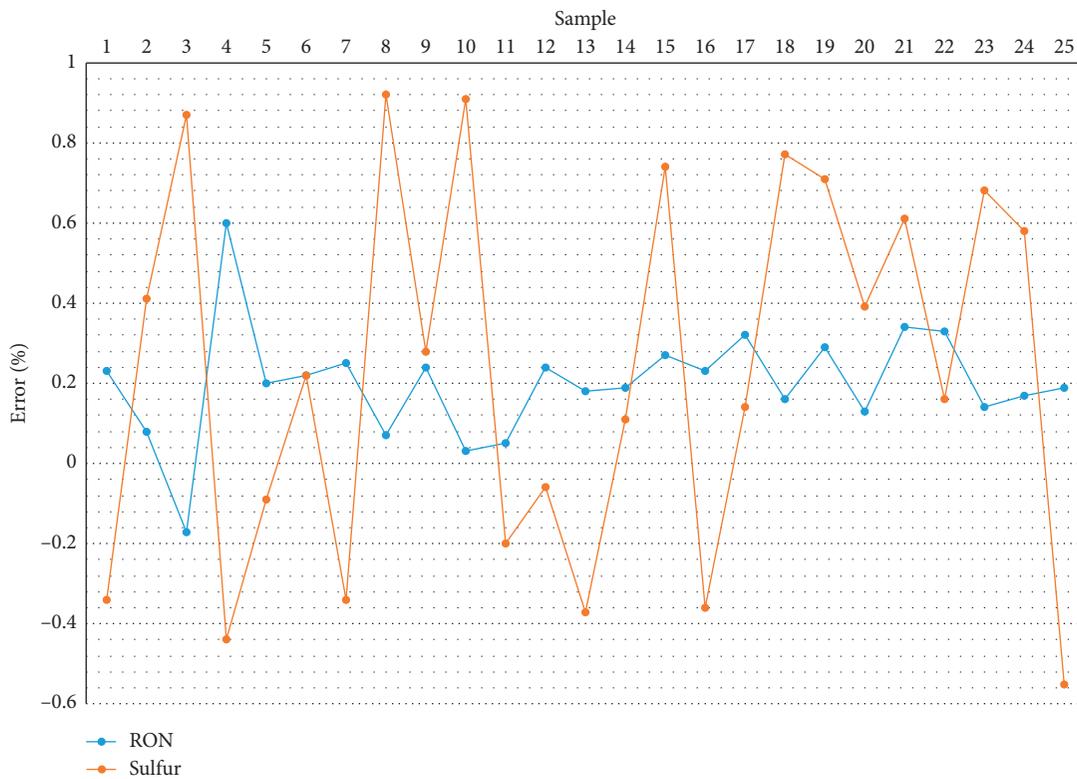


FIGURE 9: Error percentage between the predicted value and the true value.

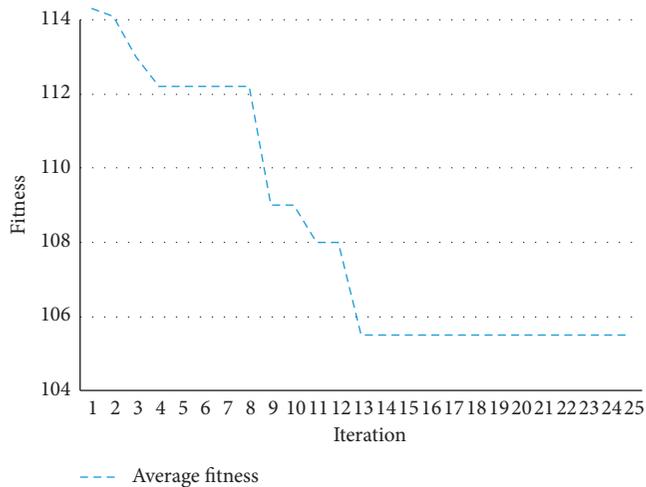


FIGURE 10: The fitness curve.

7. Conclusion and Discussion

7.1. Conclusion. Automobile traffic is a major source of pollution. Based on data analysis, this paper proposes a new research octane number extraction method to control the harm of automobile pollution from the source. This research result shows the following:

- (1) There are 27 main variables artificial remained after dimension reduction by the locally linear embedding method and calculating correlation coefficients. They are D105 temperature, 1.0 MPa steam inlet temperature, deaerated water inlet flow rate, purify air inlet flow rate, 3# catalytic gasoline intake flow rate, the oxygen content of regenerated flue gas, D204 liquid level, D203 top outlet pipe temperature, D122 outlet pipe temperature, D105 upper jumper loose wind flow rate, regenerator top/regenerator receiver differential pressure, cold nitrogen filter ME-114 differential pressure, closed hopper N2 filter outlet gas flow rate, lock hopper H2 filter outlet gas flow rate, R-102 bottom nozzle differential pressure, Me-108 filter differential pressure, K-101a left exhaust temperature, HV2533 hand operator, E-205 inlet pipe temperature, S-ZORB.AT-0003, S-ZORB.AT-0010, degassing of gasoline products, hydrocracking light naphtha intake accumulated flow, 8.0 MPa hydrogen to circulating hydrogen compressor intake flow rate, 8.0 MPa hydrogen to back blow hydrogen compressor outlet accumulated flow, the octane number of raw materials, and sulfur content of raw gasoline.
- (2) After employing the GA-BP neural network to obtain the prediction model of gasoline octane number loss, the optimal operating conditions of the main operating variables are achieved (see Table 3) under the condition that product sulfur content is less than

5 $\mu\text{g/g}$. With the optimal operating conditions can we get the optimal effect of reducing octane loss.

The realization of the optimal operating condition calculated above is conducive to the following aspects:

- (1) Improving the quality of refined oil products of chemical enterprises. By debugging the above operating points and controlling the corresponding operating conditions such as temperature, flow, or pressure difference, enterprises can improve the quality of the finished oil.
- (2) Saving the economic cost of chemical enterprises. By optimizing operating conditions to reduce the loss of octane number, the economic benefits of enterprises can be improved.
- (3) Protecting the atmospheric environment. By optimizing the operating conditions to make the refined oil meet the national standards, it is beneficial to reduce sulfide emissions and reduce air pollution.

7.2. Future Outlook. In the future, research can be carried out in the following directions.

Firstly, combining the catalytic cracking process and principles based on dimensionality reduction can better select the variables of the structural model.

Secondly, the neural network has relatively high requirements for data, but the quality and accuracy of the data in the example are not enough, so the model has a low degree of fit. In the future, more data can be collected or the data preprocessing process can be optimized to obtain data with fewer empty values and outliers.

Thirdly, with the development of data analysis technology, other intelligent machine learning algorithms can be used in the future, and comparative analysis can be performed to select better methods [47, 48].

Finally, for the automobile industry, research can be developed on how to reduce the energy consumption of automobile engines. Reducing energy consumption by improving engine performance is also an important way to reduce gasoline consumption and air pollution.

Appendix

A. Locally Linear Embedding

Locally linear embedding is [36, 37] as follows.

A.1. Locally Linear Range. Firstly, the locally linear range is solved by using the K-nearest neighbor principle. Because of locally linearity, each data point x_i can be represented by a linear combination of its nearest neighbor data points. That is, $x_i = \sum_{j=1}^k w_{ji} x_{ji}$, and $N_i = knn(x_i, k)$, $N_i = [x_{1i}, \dots, x_{ki}]$, where w_i is $k \times 1$ column vector, w_{ji} is row j of w_i ,

x_{ji} is the j th nearest neighbor to x_i , ($1 \leq j \leq k$), $w_i = [w_{1i}, \dots, w_{ki}]^T$ and $x_i = [x_{1i}, \dots, x_{Di}]^T$, where D is the dimension of x_i

Solving the weight coefficient matrix means to solve the constrained optimization problem as follows:

$$\begin{aligned} \arg \min & \sum_{i=1}^N x_i - \sum_{j=1}^k w_{ji} x_{ji}^2, \\ \text{s.t.} & \sum_{j=1}^k w_{ji} = 1. \end{aligned} \quad (\text{A.1})$$

Thus, the expression of the weight coefficient matrix can be deduced as follows:

$$\begin{aligned} \phi(w) &= \sum_{i=1}^N x_i - \sum_{j=1}^k w_{ji} x_{ji}^2 = \sum_{i=1}^N x_i - \sum_{j=1}^k w_{ji} (x_i - x_{ji})^2, \\ &= \sum_{i=1}^N X_i - \sum_{j=1}^k w_i (X_i - N_i)^2, \\ X_i &= [x_i, \dots, x_i], N_i = [x_{1i}, \dots, x_{ki}], \\ &= \sum_{i=1}^N w_i^T (X_i - N)^T (X_i - N) w_i. \end{aligned} \quad (\text{A.2})$$

Then, view S_i as local covariance matrix, $S_i = (X_i - N_i)^T (X_i - N_i)$. Eq. (A.2) can be viewed as follows:

$$\phi(w) = \sum_{i=1}^N w_i^T S_i w_i. \quad (\text{A.3})$$

Use the Lagrange multiplier,

$$L(w_i) = w_i^T S_i w_i + l(w_i^T \mathbf{1}_k - 1), \quad (\text{A.4})$$

where $\mathbf{1}_k$ is $k \times 1$ column vector with entries of 1.

Take derivative of Eq. (A.4) as follows:

$$\frac{\partial L(w_i)}{\partial w_i} = 2S_i w_i + l \mathbf{1}_k = 0, \quad (\text{A.5})$$

$$w_i = \frac{S_i^{-1} \mathbf{1}_k}{\mathbf{1}_k^T S_i^{-1} \mathbf{1}_k}.$$

A.2. Low-Dimensional Representation. The low-dimensional representation should have the same locally geometric property, so the same linear representation expression is used to finally form the quadratic form. Therefore, it is mapped to a low-dimensional space to solve the following constrained optimization problem:

$$\begin{aligned} \arg \min_Y & y(Y) = \sum_{i=1}^N y_i - \sum_{j=1}^k w_{ji} y_{ji}^2, \\ \text{s.t.} & \sum_{j=1}^k y_i y_j^2 = N I_{d \times d}. \end{aligned} \quad (\text{A.6})$$

The output result is the $d \times N$ matrix $Y = [y_1, y_2, \dots, y_N]$, which is composed of low-dimensional space vectors. Use sparse matrix W to represent w as follows:

$$\begin{aligned} \sum_{j=1}^N w_{ji} y_{ji} &= \sum_{j=1}^k w_{ji} y_{ji} = Y W_i, \\ \begin{cases} W_{ji} = w_{ji}, & j \text{ is the nearest neighbor to } i, \\ W_{ji} = 0, & \text{others,} \end{cases} \end{aligned} \quad (\text{A.7})$$

where W_i is the i column of W ($N \times N$), I_i is the i column of unit matrix I ($N \times N$), and y_i is the i column of Y .

So,

$$y(Y) = \sum_{i=1}^N Y (I_i - W_i)^2 = \text{tr}[Y(I - W)(I - W)^T Y^T]. \quad (\text{A.8})$$

Make $M = (I - W)(I - W)^T$ and use the Lagrange multiplier again as follows:

$$L(Y) = Y M Y^T + l(Y Y^T - N I). \quad (\text{A.9})$$

Take derivative of Eq. (A.9) as follows:

$$\frac{\partial L}{\partial Y} = 2M Y^T + 2l Y^T = 0. \quad (\text{A.10})$$

A.3. Take Eigenvectors. It can be seen that Y is actually the matrix composed of the Eigen vectors of M . To reduce the data to D dimension, we only need to take the minimum d eigenvectors corresponding to nonzero Eigen values of M . Generally, the first minimum Eigen value is close to 0, so we abandon it. Finally, we take the eigenvectors corresponding to the previous $[2, d + 1]$ Eigen values from the smallest to the largest.

B. A BP Neural Network Based on Genetic Algorithm

BP neural network based on the genetic algorithm is [38–45] as follows.

B.1. Initialization Network. Firstly, determine the number of network nodes, training set, and test set, and the data are normalized. Then, the individual in the genetic algorithm is initialized with real coding. Each individual is composed of four parts, namely, the weight matrix W_a between the input layer and the hidden layer, the threshold vector q_a of the hidden layer, the weight matrix W_b between the hidden layer and the output layer, and the threshold vector q_b of the output layer, forming a definite neural network.

B.2. BP Neural Network. (1) Forward propagation:

The input layer is x_1, x_2, \dots, x_n , with the weight matrix W_a between the input layer and the hidden layer. Use the linear weighted sum method to obtain the net input of the i neuron of the hidden layer which is $\text{Netin}_i = \sum_{j=1}^n x_j W_{a_{ij}}$. The net input is compared with the threshold vector q_a of the hidden layer, and then the neuron output can be obtained by activation function. Sigmoid function $f(x) = 1/1 + \exp(-x)$ is selected here as the activation function to realize signal transformation.

Therefore, the output of i neuron of the hidden layer is as follows:

$$y_i = f(\text{Netin}_i - q_{a_i}). \quad (\text{B.1})$$

Then, with the weight matrix W_b between the hidden layer and the output layer, use the same linear weighted sum method to obtain the net input of the i neuron of the output layer which is $\text{Outnetin}_i = \sum_{j=1}^n y_j W_{b_{ij}}$. The net input is compared with the threshold vector q_b of the output layer likewise, and then the i neuron output can be obtained by inverse function of Sigmoid as follows:

$$p_i = f^{-1}(\text{Outnetin}_i - q_{b_i}). \quad (\text{B.2})$$

The real output is z_1, z_2, \dots, z_m , and the least square error of the k th prediction result is as follows:

$$E_K = \frac{1}{2} \sum_{i=1}^m (p_i - z_i). \quad (\text{B.3})$$

(2) Backward propagation:

The purpose of backward propagation is to reduce the prediction error E_K and optimize the neural network. Here, the gradient descent method is used to update and reduce the parameters. The parameter adjustment formula is as follows:

$$P = -h \left(\frac{\partial E_K}{\partial P} \right), \quad (\text{B.4})$$

where h is the leaning rate.

After adjusting the aforementioned parameters, the neural network is optimized by repeating iteration and adjustment.

(3) Termination conditions:

When one of the following termination conditions is reached, the iteration is stopped and a neural network is obtained.

- (1) The minimum error threshold E_{\min} is reached
- (2) The maximum number of iterations G_{\max} is reached

B.3. Calculate Fitness. After running the neural network with the training set data, the prediction system output was

obtained for each individual and the fitness of the individual F_i is defined as the error E_i between the predicted output and the real output, and the formula is as follows:

$$F_i = E_i = \frac{1}{2} \sum_{i=1}^n (p_i - z_i), \quad (\text{B.5})$$

where n is the number of output nodes; p_i is the predicted output of the i node, and z_i is the real output of the i node.

B.4. Selection, Crossover, and Mutation. (1) Selection:

Selection means to retain the high-quality individual, while eliminating the poor quality individual. The Roulette algorithm is chosen for selection operation. Firstly, calculate the probability that an individual will be selected:

$$p_i = \frac{f_i \cdot k}{\sum_{i=1}^N f_i}, \quad (\text{B.6})$$

where $f_i = (k/P_i)$ is the reciprocal of the fitness, which is positively correlated with adaptability; k is the coefficient, which is set at 1; and N is the population size.

Then, the interval $[0, 1]$ is divided into N intervals, the length of which is the same as the probability of an individual will be selected. The location interval is determined by generating random numbers so that corresponding individuals could be selected to form a new population.

(2) Cross:

Cross is individuals of new populations exchange gene fragments to generate new individuals. Since individuals are real coding, the real cross method is adopted here for crossover operation.

Firstly, the individual in the group is judged to carry out the cross with the set crossover probability. The formula of cross between the m individual and the n individual at the j intersection position is as follows:

$$\begin{cases} a_{mj} = a_{mj}(1-b) + a_{nj}b, \\ a_{nj} = a_{nj}(1-b) + a_{mj}b, \end{cases} \quad (\text{B.7})$$

where b is the random number in $[0, 1]$.

(3) Mutation:

Mutation is changes in the individual's own gene value. Since individuals are real coding, real value mutation is adopted here for mutation operation.

Firstly, the individual in the group is judged to carry out the mutation with the set mutation probability. Then, the j mutation location of the i individual was selected for the mutation operation. The formula is as follows:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) \times f(g), & r > 0.5, \\ a_{ij} + (a_{\min} - a_{ij}) \times f(g), & r \leq 0.5, \end{cases} \quad (\text{B.8})$$

where a_{\max} is the upper bound of a_{ij} , a_{\min} is the lower bound of a_{ij} , r is the random number in $[0, 1]$, $f(g)$ is the mutation formula: $f(g) = r_2(1 - (g/G_{\max}))^2$, r_2 is the random number in $[0, 1]$, g is the current iteration number, and G_{\max} is the maximum number of iterations.

When one of the following termination conditions is reached, the iteration is stopped, and the optimal initial weight and threshold are obtained.

- (1) The fitness of the optimal individual and the fitness of the population stopped rising
- (2) The maximum number of iterations G_{\max} is reached

B.5. Optimal Prediction Model Building. When the optimal initial weight and threshold are obtained by the genetic algorithm above, they are given to BP neural network for the optimal network forming. When one of termination conditions is reached, iteration is stopped, and the optimal neural network is obtained.

Data Availability

Historical accumulated data of a petrochemical company's S-Zorb device are collected, including operating variable data, raw material data, product data, and catalyst data. After adjusting the data frequency difference, 325 samples are formed.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Social and Scientific Fund Program of China (17BGL142 and 18ZDA052) and Natural Science Foundation of China (91546117 and 71904117).

Supplementary Materials

(1) S-Table1: the historical accumulated data of the catalytic cracking gasoline S-Zorb unit. There are 325 samples of operating variable data, raw material data, product data, and catalyst data observed in different time intervals within 2017-2020. (2) S-Table2: the certain operating range and minimum adjustment range of operating variables. (*Supplementary Materials*)

References

- [1] OICA, *Registrations or Sales of New Vehicles*, Organisation Internationale des Constructeurs d'Automobiles, Paris, France, 2019.
- [2] IEA, *Oil Information: Database Documentation*, International Energy Agency, Paris, France, 2020.
- [3] R. A. Field, J. N. Lester, S. O. Baek, M. E. Goldstone, P. W. Kirk, and R. Perry, "A review of atmospheric polycyclic aromatic hydrocarbons: sources, fate and behavior," *Water Air & Soil Pollution*, vol. 60, no. 3-4, pp. 279-300, 1991.
- [4] Health Effects Institute, "Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects," *Environment*, vol. 131, pp. 409-444, 2010.
- [5] N. R. Khalili, P. A. Scheff, and T. M. Holsen, "PAH source fingerprints for coke ovens, diesel and, gasoline engines, highway tunnels, and wood combustion emissions," *Atmospheric Environment*, vol. 29, no. 4, pp. 533-542, 1995.
- [6] R. K. Larsen and J. E. Baker, "Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: a comparison of three methods," *Environmental Science & Technology*, vol. 37, no. 9, pp. 1873-1881, 2003.
- [7] M. Riedl and D. Diaz-Sanchez, "Biology of diesel exhaust effects on respiratory function," *Journal of Allergy and Clinical Immunology*, vol. 115, no. 2, pp. 221-228, 2005.
- [8] IEA, *World Energy Balances: Database Documentation*, International Energy Agency, Paris, France, 2020.
- [9] F. Ouyang, W. Fang, J. Tang, and H. Jiang, "Optimizing product distribution of mip process using BP neural network," *Petroleum Processing and Petrochemicals*, vol. 47, no. 5, pp. 95-100, 2016.
- [10] E. Paranghooshi, M. T. Sadeghi, and S. Shafiei, *Prediction of Octane Number for Gasoline Blends Using Artificial Neural Networks and Genetic Algorithms*, ACS Publications, Washington, DC, USA, 2009.
- [11] X. J. Qin and Z. H. Chen, "The design of the neural network model for predicting gasoline RON," *Control and Decision*, vol. 14, no. 2, pp. 151-155, 1999.
- [12] J. Tian, X. Du, Y. Zheng, Y. Li, and T. Jing, "Prediction of sulfur content in hydrodesulfurization diesel oil based on the PSO-BP neural network," *Petrochemical Technology*, vol. 46, no. 1, pp. 62-67, 2017.
- [13] Y. L. Yang, X. D. Zhang, and H. Jiang, "BP network based on GA-BP optimization algorithm and application in gasoline blending octane number modeling," *Microcomputer Information*, vol. 22, no. 11, pp. 276-278, 2006.
- [14] S. Brunet, D. Mey, G. Pérot, C. Bouchy, and F. Diehl, "On the hydrodesulfurization of FCC gasoline: a review," *Applied Catalysis A: General*, vol. 278, no. 2, pp. 143-172, 2005.
- [15] J. A. Valla, A. A. Lappas, I. A. Vasalos, C. W. Kuehler, and N. J. Guddé, "Feed and process effects on the in situ reduction of sulfur in FCC gasoline," *Applied Catalysis A General*, vol. 276, no. 1-2, pp. 75-87, 2004.
- [16] Y. N. Yang, Y. Ren, A. G. Mao, and H. P. Tian, "Analysis of technical factors affecting RON of FCC gasoline," *Petroleum Refinery Engineering*, vol. 1, no. 6, pp. 32-35, 2019.
- [17] S. M. Jacob, B. Gross, S. E. Voltz, and V. W. Weekman, "A lumping and reaction scheme for catalytic cracking," *AIChE Journal*, vol. 22, no. 4, pp. 701-713, 1976.
- [18] L. G. Lin, G. Wang, H. M. Qu et al., "Pervaporation performance of cross-linked polyethylene glycol membranes for deep desulfurization of FCC gasoline," *Journal of Membrane Science*, vol. 280, no. 1-2, pp. 651-658, 2006.
- [19] C. Song, "An overview of new approaches to deep desulfurization for ultra-clean gasoline, diesel fuel and jet fuel," *Catalysis Today*, vol. 86, no. 1-4, pp. 211-263, 2003.
- [20] M. A. B. Siddiqui and A. M. Aitani, "FCC gasoline sulfur reduction by additives: a review," *Petroleum Science and Technology*, vol. 25, no. 3, pp. 299-313, 2007.
- [21] J. A. Salazar, L. M. Cabrera, E. Palmisano, W. J. Garcia, and R. B. Solari, "Process for producing reformulated gasoline by reducing sulfur, nitrogen and olefin," U. S. Patent 5770047, 1998.

- [22] M. F. Li, G. F. Xia, Y. Chu, and Y. J. Hu, "Development of RSDS-Icatalyst for selective hydrodesulfurization of FCC gasoline," *Petroleum Processing and Petrochemicals*, vol. 34, no. 7, pp. 1–4, 2003.
- [23] H. Li, W. Shan, S. Shen et al., "Production of a gasoline blending component with high-octane and low sulfur from coal tar light oil over sulfided CoMoP/η-Al₂O₃," *Journal of Cleaner Production*, vol. 228, no. 10, pp. 965–973, 2019.
- [24] S. Ayoub and V. Masoud, "Design and optimization of hydrodesulfurization process for liquefied petroleum gases," *Journal of Cleaner Production*, vol. 220, no. 20, pp. 1255–1264, 2019.
- [25] N. Hasheminejad, H. Tavakol, and W. Salvenmoser, "Preparation of gold-decorated simple and sulfur-doped carbon spheres for desulfurization of fuel," *Journal of Cleaner Production*, vol. 264, no. 10, Article ID 121684, 2020.
- [26] X. Qin, J. Liu, C. Wang et al., "Molecular level analysis on performance of diameter expanding reactor to improve gasoline quality in FCC process," *Fuel*, vol. 290, Article ID 119978, 2021.
- [27] M. G. Li, Y. Y. Xu, J. Men et al., "Hybrid variable selection strategy coupled with random forest (RF) for quantitative analysis of methanol in methanol-gasoline via raman spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 251, no. 1, Article ID 119430, 2021.
- [28] G. F. Wei, "Study on gas recognition method with principal component analysis and back-propagation neural network," *Journal of Translucation Technology*, vol. 12, no. 04, pp. 292–298, 2001.
- [29] H. X. Cheng and F. Yi, "Light gasoline etherification predictive control with BP neural network model," *Automation in Petro-Chemical Industry*, vol. 48, no. 6, pp. 40–42, 2012.
- [30] Z. Zhang, Z. Li, Y. Li, and G. Li, "GA-ANN method for prediction of gasoline yield of RFCCU," *Petroleum Processing and Petrochemicals*, vol. 45, no. 7, pp. 91–96, 2014.
- [31] X. Su, H. J. Pei, Y. Y. Wu, J. S. Gao, and X. Y. Lan, "Predicting coke yield of FCC unit using genetic algorithm optimized BP neural network," *Chemical Industry and Engineering Progress*, vol. 20, no. 2, pp. 389–396, 2016.
- [32] H. Cheng, Y. Zhang, L. Kong, and X. Y. Meng, "The application of neural network PID controller to control the light gasoline etherification," *Iop Conference Series Earth and Environmental Science*, vol. 69, no. 1, Article ID 012045, 2017.
- [33] Y. Zhang, L. Zhao, F. Chen, Y. Wang, and C. Xu, "High efficiency separation of olefin from FCC naphtha: separation mechanism and universal simulation method," *AIChE Journal*, vol. 67, no. 5, Article ID e17153, 2021.
- [34] B. Foroughi, J. R. Shahrouzi, and R. Nemati, "Detection of gasoline adulteration using modified distillation curves and artificial neural network," *Chemical Engineering & Technology*, vol. 44, no. 3, pp. 527–534, 2021.
- [35] Y. Ma, Z. Yu, Y. Wang, D. Xie, and J. E., "Investigation on the influence of initial thermodynamic conditions and fuel compositions on gasoline octane number based on a data-driven approach," *Fuel*, vol. 291, Article ID 120124, 2021.
- [36] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [37] L. K. Saul and S. T. Roweis, "An introduction to locally linear embedding," *Journal of Machine Learning Research*, vol. 7, pp. 1–13, 2001.
- [38] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [39] J. Savković-Stevanović, "Neural networks for process analysis and optimization: modeling and applications," *Computers & Chemical Engineering*, vol. 18, no. 11-12, pp. 1149–1155, 1994.
- [40] C. H. Chen, "An explainable deep neural network for extracting features," *Science eLetters*, vol. 365, no. 9, p. 6452, 2019.
- [41] C. Guo, G. Liu, L. Lyu, and C.-H. Chen, "An unsupervised PM_{2.5} estimation method with different spatio-temporal resolutions based on KIDW-TCGRU," *IEEE Access*, vol. 8, pp. 190263–190276, 2020.
- [42] D. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, Addison-Wesley, Boston, MA, USA, 1989.
- [43] W. Gao, "An improved fast-convergent genetic algorithm and its performance study," *Systems Engineering and Electronics*, vol. 25, no. 11, pp. 1427–1430, 2003.
- [44] R. Chen, Y. Xu, and H. Lan, "Research on multilayer feed forward neural networks-genetic backpropagation algorithm & structure optimization strategy," *Acta Automatica Sinica*, vol. 23, no. 1, pp. 43–49, 1997.
- [45] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, no. 541, Article ID 123344, 2020.
- [46] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [47] X. Wu, Y. R. Cao, Y. Xiao, and J. Guo, "Finding of urban rainstorm and waterlogging disasters based on microblogging data and the location-routing problem model of urban emergency logistics," *Annals of Operations Research*, vol. 60, pp. 1–32, 2018.
- [48] X. Wu, Z. Xu, H. Liu, J. Guo, and L. Zhou, "What are the impacts of tropical cyclones on employment? an analysis based on meta-regression," *Weather, Climate, and Society*, vol. 11, pp. 259–275, 2019.