Simulation Method of Port Petrochemical Industry Throughput Development under the Background of Integration of Port, Industry and City

Tingting Zhou, Chen Guo*
College of Economics and Management, Yingkou Institute of Technology, Yingkou, 115014, China

***Abstract*—In order to accurately predict the changes in the throughput of port petrochemical products and facilitate the formulation of relative decisions, this paper analyzes the factors affecting the throughput of port petrochemical products in a city through the GRA method. After sorting and selection, PCA method is used for pretreatment. In the SVM algorithm, ICSO is used to obtain the best parameters and improve the prediction accuracy and efficiency. In view of the variability of future development, three development scenarios are set up to prepare for the throughput forecast of petrochemical products in a city's port. The results show that the optimization speed of ICSO algorithm is very fast. When the training iteration is 20, the best fitness value is obtained, which is 0.0572. The training effect of ICSO-SVM algorithm is good, the gap between it and the original data is small, and the overall trend is close to the original data. In the test prediction, ICSO-SVM algorithm has the best prediction effect, and its MAE, RMSE and MAPE are the smallest. The minimum MAE is 762.2, 477.0 smaller than CSO-SVM algorithm, and the latter's MAE is 1239.2. The minimum MAPE of the proposed algorithm is 1.05%, while that of CSO-SVM algorithm is 1.71%. In general, the prediction error of ICSO-SVM algorithm is smaller. After the prediction of different development scenarios, the throughput of petrochemical products in a port of a city shows an increasing trend in the next five years. This method can be applied to the development forecast of port petrochemical products and provide reference for decision-making.**

Keywords*—*Support vector machine algorithm; port throughput; chicken swarm optimization algorithm; grey correlation analysis; petrochemical products

I. INTRODUCTION

With the continuous development of the economy, the heavy chemical industries on both sides of the Yangtze River are facing deep adjustment. Due to the low economic efficiency of the chemical siege movement, the waste of land resources and the pollution of the ecological environment have resulted in the lack of impetus for the integration and development of port clusters, petrochemical industry clusters and urban clusters, that is, it is difficult for the three to coordinate [1-3]. In order to promote the coordinated development of port, industry and city, it is necessary to analyze the development of each integrated part in depth. Some scholars choose Support Vector Machine (SVM) to predict the port throughput when studying it. In order to explore the impact of market indicators, they use this indicator as the input of the prediction model. Through experimental tests, the model has achieved good prediction results and can predict port throughput [4]. To predict the power in the Kanto region, some scholars choose SVM algorithm to build relevant multi-network configuration predictors. After testing, the accuracy of prediction has been improved, and the test effect is good [5]. Therefore, when studying the throughput of port petrochemical products, SVM algorithm is selected as the prediction algorithm. In the analysis and processing of the factors affecting the throughput of port petrochemical products, the grey correlation analysis (GRA) and principal component analysis (PCA) are used for correlation analysis and processing to predict the throughput of port petrochemical products more accurately. By predicting the throughput of petrochemical products in ports, it is beneficial for ports to grasp the direction of business development and actively promote the rational layout of port logistics networks. The study was divided into four parts. The second part is a literature review, which introduces the research of domestic and foreign scholars on port development and port throughput, the good performance of SVM algorithm in forecasting and the relevant application research of PCA and GRA method in data analysis, among which these three methods are also suitable for port throughput forecasting research. In the third part, PCA method and GRA method are proposed to deal with the factors affecting the throughput of petrochemical products at port. ICSO-SVM algorithm is used to predict the throughput of petrochemical products at port. The fourth part carries on the empirical analysis of port petrochemical product throughput prediction, and the results show the superiority of ICSO-SVM algorithm.

II. RELATED WORK

To promote the development of the integration of port, industry and city, the transformation of petrochemical industry is necessary. Relevant parties need to adjust their industrial layout, strengthen the internal power of coordinated development, and protect the ecological environment while coordinating development. During the transformation of the petrochemical industry, some scholars simulated the development of the transportation volume of the port petrochemical industry to understand its development trend and make better decisions. In the face of the transportation problem of petrochemical products, An H and others adopted the relevant comprehensive short-term scheduling model for scheduling. They used heuristic algorithms. The data set test confirmed the superiority and low cost of this method [6]. Zhang selected the port and shipping industry of Ningbo as the research object to study its development strategy. Taking the spatial scale as the starting point, they put forward their relevant strategic objectives, analyzed the relevant paths, and finally gave the relevant strategic integration model [7]. Wang and others analyzed the national economy and the role of port industry in it. They focus on their input and output, as well as the relevant time evolution process. They summarized relevant policy applications [8]. Ngoc et al. studied container port throughput management and applied control theory and chaos analysis. They optimize the port operation through the theory of sliding membrane control. The test verifies the effectiveness of this method [9].

Zhang et al., faced with the problem of carbon price prediction, chose the least squares SVM to predict it. When processing raw data, they choose the empirical model. The results show that the method is effective and feasible [10]. Praveena et al. detected epileptic seizures and processed the collected data through PCA dimensionality reduction by intelligent means. They use SVM to classify them. After verification, the application effect of this method is good [11]. Song et al. carried out earthquake early warning P-wave prediction on the basis of SVM. After model training, the test error of this method is small [12]. In order to predict PM2.5 of air pollution, Lai X et al. carried out relevant prediction by improving SVM algorithm on the basis of feature selection. The results confirmed the availability of this method [13]. Huang J et al. faced the problem of analyzing the factors affecting the calcination temperature of a vertical furnace, and based on orthogonal design, chose the GRA method to quantify the significant factors involved. According to the analysis results, the influence of volatile matter content is significant and is a key factor [14]. Chen C et al. used the PCA method to reduce the dimensionality of the relevant influencing factors in order to predict the phosphorus content at the endpoint of the Condi electric furnace, and input the processed results into the extreme randomization tree. The results show that the prediction effect is good [15]. Luo S et al. chose to improve the SVM algorithm when facing the problem of predicting the thermal state inside the blast furnace skull. After testing on the dataset, this method has high prediction accuracy [16].

To sum up, in the study of port throughput, there are relatively few researches on its cargo, which mainly focus on the port itself and analyze the development of the port. The researches on cargo throughput are not deep enough. In order to dig into the key factors affecting the development and change of the throughput of petrochemical products in port and obtain more accurate throughput prediction results, this study takes the throughput of petrochemical products in port as the research object and studies its development trend. Through GRA method and PCA method, the key influencing factors are dug out. Considering the good effect of SVM algorithm in prediction, the improved SVM algorithm is used to carry out relevant research, which makes the prediction technology be innovatively improved and the prediction accuracy and generalization ability of the prediction model be enhanced.

III. Simulation of Port Petrochemical Industry Traffic Volume Development based on ICSO-SVM Algorithm under the Background of Port Industry and City Integration

A. *Treatment of Factors affecting Port Petrochemical Product Throughput based on GRA and PCA Methods*

In the continuous economic development, the integration of port and industry is an important way to develop regional economic, which is conducive to the transformation of petrochemical industry and accelerating the integration process. To understand the development trend of the petrochemical industry in the development of transportation volume, the article selects the port petrochemical product throughput as the research object and analyzes it. The change trend of the throughput of petrochemical products in a city's port in recent years is shown in Fig. 1.

In Fig. 1, the throughput of port petrochemical products in 2020 was lower than that in 2019. In 2021, the throughput of petrochemical products in the port returned to normal and increased to a certain extent, surpassing 2019. Although there were gaps in the throughput of petrochemical products in ports in different years, there was an overall growth trend. To study the causes of the changes in the throughput of port petrochemical products, this paper analyzes the influencing factors. Through consulting information and data acquisition, it is found that the throughput of port petrochemical products is affected by many factors. Based on these factors, the paper constructs a pre-selection index system of relevant influencing factors, as shown in Fig. 2.

In Fig. 2, the indicator system includes four primary indicators, namely economic and trade level (EATL), regional development vitality (RDV), population and employment level (PAEL), port infrastructure conditions (PIC), and 14 secondary indicators. In order to understand the correlation between these indicators and the throughput of port petrochemical products, the article selects the GRA method to analyze them. The greater the correlation, the stronger the corresponding correlation. In the GRA method, first set the original sequence $X'_i = [x'_i(1), x'_i(2), \cdots, x'_i(n)]$, $i = 0,1,2,\cdots,l$. The length of the number sequence is set to n, and there are $l+1$ index number sequences collected [17-19]. The difference sequence is solved for the initial value of each sequence, and its calculation formula is shown in Eq. (1).

$$\Delta_i(k) = \left|x'_0(k) - x'_i(k)\right|, \Delta_i = (\Delta_i(1), \Delta_i(2), \cdots \Delta_i(n)) \quad (1)$$

In formula (1), Δ_i represents the difference sequence and k represents the sequence number. Solve the maximum difference and minimum difference between the two poles, as shown in Formula (2).

$$M = \max_l \max_k \Delta_i(k), M = \min_l \min_k \Delta_i(k) \quad (2)$$

In formula (2), M represents the maximum difference between the two poles, and m represents the minimum difference between the two poles. Solve the correlation coefficient $\gamma_{0i}(k)$ and calculate the correlation degree. The calculation formula is shown in formula (3).

$$\gamma_{0i} = \frac{1}{n}\sum_{k=1}^{n}\gamma_{0i}(k), i = 1, 2, \cdots l \qquad (3)$$

In formula (3), γ_{0i} represents the degree of correlation. The software used for correlation calculation is MATLAB. Sort the relevant results according to the order from the largest to the smallest. According to the ranking of the correlation degree obtained, 13 indicators are selected as the prediction indicators required by the article. These indicators are in the top 13 in the ranking of correlation degree and have high correlation with the throughput of port petrochemical products. Among them, the top three indicators are GDP, total import and export of petrochemical products, and coastal berths. In particular, the correlation value corresponding to GDP is the largest. In order to better reflect the impact factors of port throughput, 13 indicators are included in the prediction model. However, the dimension of data is too large to have a certain impact on the prediction effect of SVM. Therefore, PCA method is adopted to reduce data redundancy and reduce the complexity of the problem [20-22]. It is a linear dimensionality reduction algorithm with practical significance, which can perform orthogonal transformation on high-dimensional data. The transformed data maintains the basic information of the original data. After the change of linearly related variables, the new variable formed has linear independence, which is also called principal component.

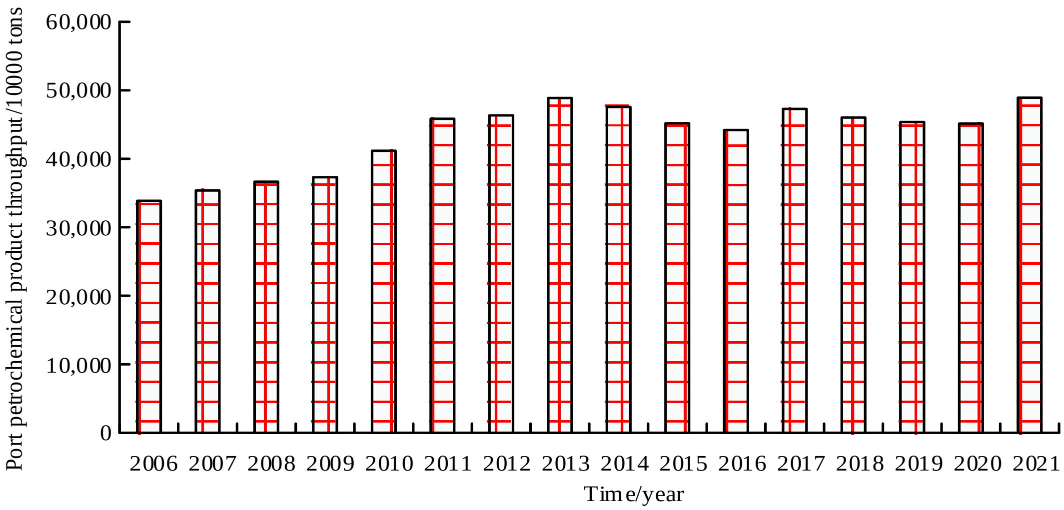


Fig. 1. Change trend of petrochemical product throughput in a port.

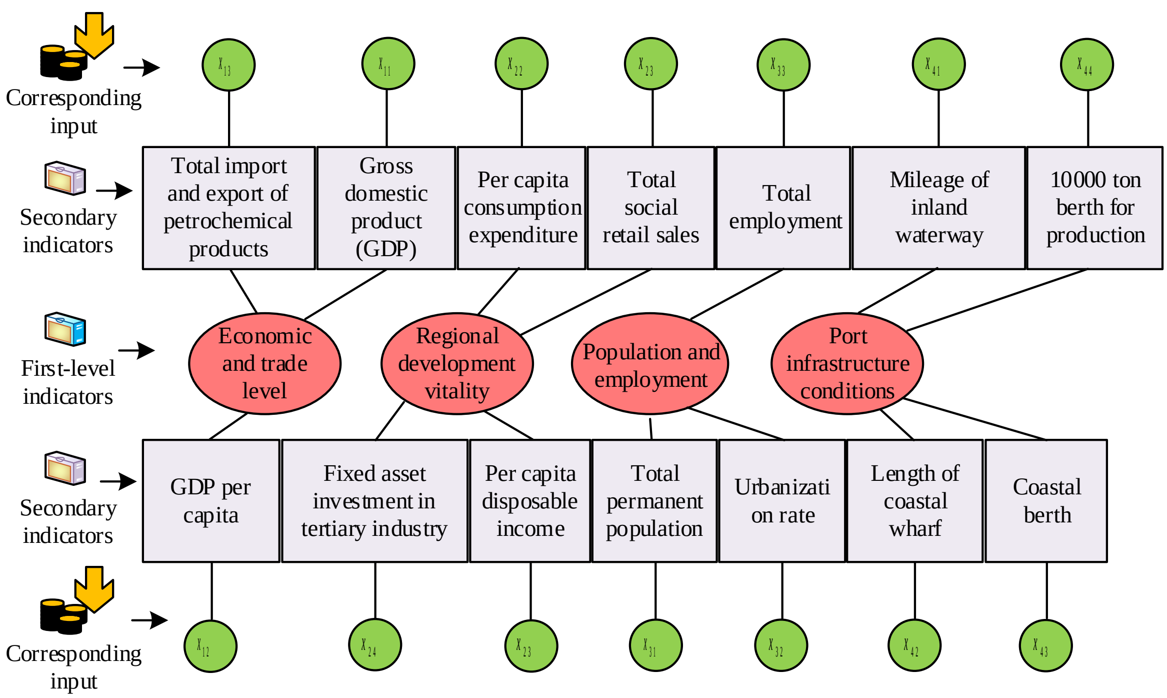


Fig. 2. Pre-selection index system of influencing factors.

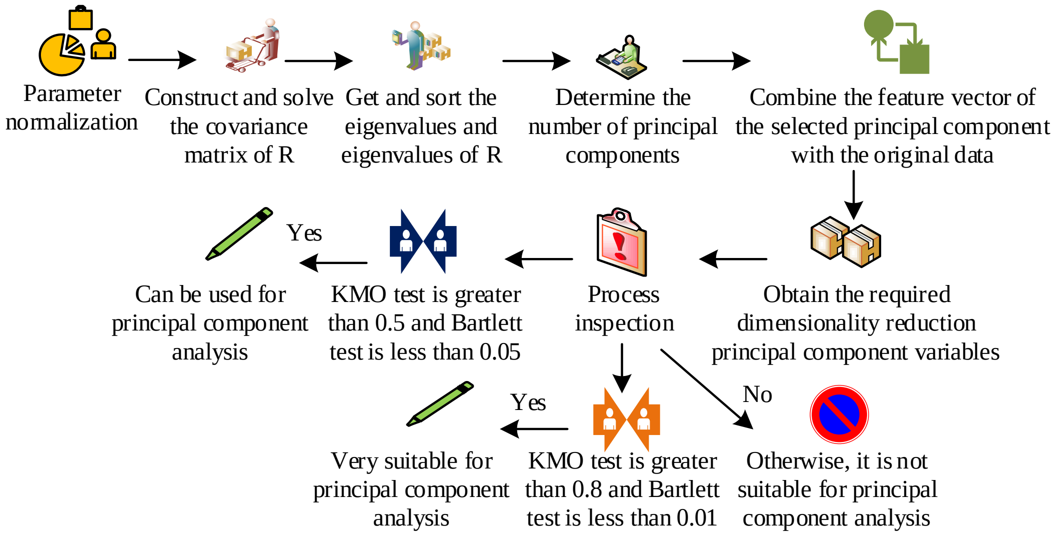


Fig. 3. Relevant process of PCA method.

PAC method first normalizes all parameters to form a normalized matrix. Secondly, it obtains the eigenvalues and eigenvalues of R by constructing and solving the covariance matrix of R, and ranks them according to the variance of principal components. The number of principal components is determined by their contribution degree and cumulative variance contribution rate, and the eigenvalues are taken as the judgment criteria, which need to exceed 1, and the contribution rate need to be no less than 85%. The feature vector of the selected principal component is combined with the original data to obtain the required dimension-reduced principal component variable. Then, there are two test methods in dimension reduction, namely, Kaiser-Meyer-Olkin (KMO) test and Bartlett test. If the value of the former test method is greater than 0.5 and the value of the latter test method is less than 0.05, the principal component analysis can be performed. If the value of the former test method is greater than 0.8, and the value of the latter test method is less than 0.01, it is particularly suitable for principal component analysis. The relevant process of PCA method is shown in Fig. 3.

After dimension reduction by PCA method, the principal components are obtained, and the correlation between the factors is eliminated, which maximizes the original main information. This method is simple, easy to implement, without parameter constraints, and has good objectivity. It can not only simplify the subsequent data processing, but also help the SVM model have better robustness. The PCA method is used to reduce the dimension of the selected 13 indicators. To better save the relevant information of the original data, the PCA method is used to process the secondary indicators under a primary indicator, and then the relevant secondary indicators of the remaining primary indicators are processed according to this method. The software used for PCA analysis is SPSS software.

B. Application of ICSO-SVM Algorithm in Throughput Prediction of Port Petrochemical Products

According to the relevant data characteristics of port petrochemical product throughput, forecasting and analysis by SVM algorithm can effectively overcome the problems of other models. Problems include insufficient sample size and poor model stability. This method is good at processing a kind of nonlinear data, which is characterized by small samples and high dimensions. The SVM algorithm is to find an optimal hyperplane. On both sides of the plane, the samples are farthest away from it, and the classification effect and robustness of the algorithm are also better [23-24]. In the distance between sample point and hyperplane, the relevant calculation formula is shown in formula (4).

$$d = \frac{\left\| w^T x + b \right\|}{\left\| w \right\|} \tag{4}$$

In formula (4), x_i represents sample point and d represents sample interval. The normal vector is expressed as w, b means parameter. To further describe the sample interval, some training samples closest to the hyperplane on the hyperplane are found, and a formula that meets the conditions is given as formula (5).

$$\begin{cases} w^T x_i + b = 1, y_i = 1 \\ w^T x_i + b = -1, y_i = -1 \end{cases} \tag{5}$$

In formula (5), y_i represents the sample point. According to Formula (4) and Formula (5), the sum of the distance between the two support vectors and the hyperplane is $\frac{2}{\left\| w \right\|^2}$. When this value is maximum, the corresponding sample discrimination is maximum. Analyze the nature of the sample and the degree to which it can be separated, and use the kernel function SVM as its learning method. Support vector machine regression (SVR) is widely used in classifiers. When constructing the regression function $f(x)$, set the loss boundary as ε. When the distance between $f(x)$ and sample is less than ε, an isolation band with the width of 2ε is formed. When the sample is in the isolation zone, the prediction is correct. After the kernel function is applied, the relevant objective function is shown in equation (6).

$$\min \frac{\|w\|^2}{2} + C\sum_{i=1}^{N} \ell_\varepsilon (f(x_i) - y_i) \tag{6}$$

In formula (6), the penalty factor is expressed as C, and the ε insensitive loss function is expressed as l_ε. The relevant expression is shown in Eq. (7).

$$\ell_\varepsilon = \begin{cases} 0, if\ |z| \le \varepsilon \\ |z| - \varepsilon, else \end{cases} \tag{7}$$

In formula (7), z represents the parameter. By adding relaxation variables δ_i and $\hat{\delta}_i$, the SVR problem is transformed. Combining Lagrange multipliers α_i and $\hat{\alpha}_i$, the relevant dual problem can be obtained, and then the optimal solution of SVR problem is shown in Eq. (8).

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \hat{\alpha}_i) K(x_i, x) + b \tag{8}$$

In formula (8), K represents the parameter. In SVM, C and kernel function width σ have a greater impact on its prediction effect. In view of this situation, radial basis function (RBF) is selected as the kernel function in this paper because of its strong generalization ability. In addition, to obtain the best parameters, the paper selects CSO algorithm for parameter optimization. Because of its poor processing effect on complex optimization problems, this paper optimizes it through two improvement strategies and obtains the improved CSO (Improved Chicken Swarm Optimization, ICSO) algorithm. Using chaos theory, the diversity of the algorithm is strengthened, and adaptive learning strategy is introduced. In the first strategy, due to the randomness of the initial population generation of CSO algorithm, it is easy to have a local optimal solution. Adding chaotic variables can avoid this situation. This variable can be generated by a specific chaotic map, and its initial population distribution range is much larger than the random variable, and can be combined with intelligent optimization methods. Among them, tent mapping is a typical method of generating chaotic sequence similar to tent, and the formula involved is shown in formula (9).

$$x"_{i+1} = \begin{cases} x"_i / a, x" < a \\ (1 - x"_i)/(1 - a), x" \ge a \end{cases} \tag{9}$$

In formula (9), $x"_i$ and x_0 represent chaotic variables and initial variables respectively. The mapping coefficients are expressed as $a \in (0,1)$, and a group of chaotic sequences can be obtained by x_0 iterating n times. Compared with random sequence, the data distribution characteristics of the chaotic sequence of Ten-map are relatively stable, so the corresponding $x"_i$ is selected to replace the random number of CSO algorithm. In the adaptive learning strategy, the chicken's movement mode is improved, and the self-learning coefficient is added, so that the direction of its movement has a certain probability towards the cock, and the relevant calculation formula of the relevant improved strategy is shown in Formula (10) and Formula (11).

$$x_{ij}^{t+1} = w_i * x_{ij}^t + F * (x_{mj}^t - x_{ij}^t) + R * (x_{rj}^t - x_{ij}^t) \tag{10}$$

In formula (10), w_i represents the self-learning coefficient of chicks, and x_m represents the position of hens. At the time of t, the position of the individual is expressed as x^t_{ij}, the fitness of the i th hen is expressed as f_i, and the chick random following parameter is expressed as F. r represents the cock's coefficient, and R means the coefficient that follows the cock's movement, which will replace the chaotic variable.

$$w_i = \begin{cases} w_{\min} + (f_i - f_{\min}) * (w_{\max} - w_{\min}) / (f_{avg} - f_{\min}), if\ f_i \le f_{avg} \\ w_{\max} \qquad else \end{cases} \tag{11}$$

In formula (11), $w_{\min}$ and $w_{\max}$ represent the minimum and maximum weight. The maximum fitness, minimum fitness and average fitness of chicken flocks are expressed as $f_{\max}$, $f_{\min}$ and f_{avg} respectively. Combined with the improved CSO algorithm, the problem to be solved is the optimal problem on the bounded two-dimensional plane. Its goal is to find the $f_{\min}$ best individual. The individual f of the chicken flock corresponds to the mean square error of prediction, that is, to find the minimum prediction error, and then the optimal parameters of SVM can be obtained. The throughput of port petrochemical products is predicted by ICSO-SVM algorithm. The data of the first 11 years in Fig. 1 are classified as training sets, and the rest are test sets. First, data preprocessing and parameter setting are carried out. The basic model is the $\varepsilon - SVR$ model. The system default value is 0.1. Set the ICSO algorithm parameters, as shown in Fig. 4.

In Fig. 4, according to the Ten-map chaotic algorithm, all chickens form initial positions. In the iterative process, individuals search for optimization in the search space according to their own way of movement. The individual identity is updated every 10 generations. The update is based on the current f. Keep iterating until the prediction error is less than 10-4 or the maximum iteration is reached, then the operation will be stopped. For data preprocessing, Z-core standardization method is selected, as Formula (12).

$$H' = \frac{H - \bar{H}}{\sigma} \tag{12}$$

In formula (12), H', H and $\bar{H}$ are mean normalized, original and original average data respectively. Mean Absolute Error (MAE) is used to evaluate the prediction effect, as Formula (13).

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |s_t - n_t| \tag{13}$$

In formula (13), T represents the number of years, and the real original value and predicted value of the t year are s_t and n_t respectively. The evaluation index is Mean Absolute Percentage Error (MAPE), as shown in formula (14).

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{s_t - n_t}{s_t}\right| * 100\% \qquad (14)$$

The root mean square error (RMSE) as the evaluation index is shown in formula (15).

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\frac{s_t - n_t}{s_t}\right|^2} \qquad t \in 1, 2, \cdots T \quad (15)$$

After the prediction results are obtained, they are compared with other prediction methods to verify the reliability of the methods used in the article. Overall, the model building process is shown in Fig. 5.

In Fig. 5, several comparison models are applied to test and verify the performance of the methods used in the article. During the comparative experiment, the same data sets were used. These comparison methods are shown in Fig. 6.

After testing the method, it is applied to the throughput prediction of petrochemical products in a port of a city. Considering the uncertainty of future development, this paper uses scenario prediction method to predict the throughput of petrochemical products in a city's future port. First, according to the policy documents of a city and its development rules, different scenarios are set to reflect different development situations. The middle-aged change rate of different scenarios is set to obtain the corresponding future value. These values are entered into the model. Through the trained model, the throughput under different scenarios is predicted and the relevant outputs are obtained.

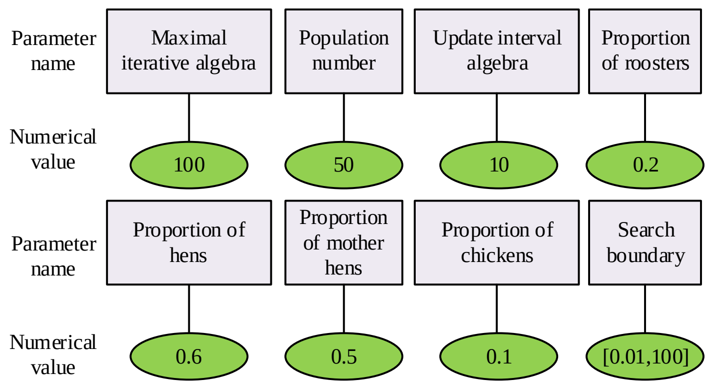


Fig. 4. ICSO algorithm parameters.

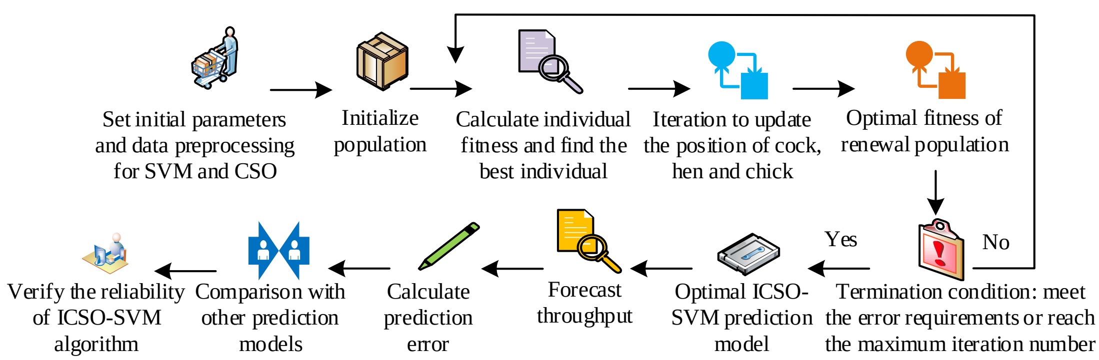


Fig. 5. Model building process.

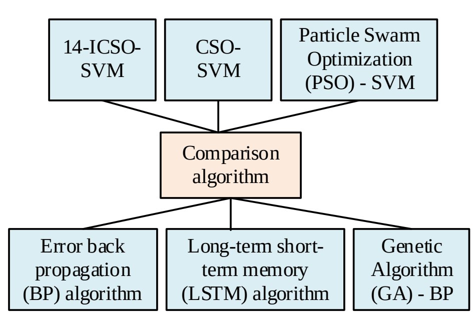


Fig. 6. Relevant comparison methods.

IV. Prediction and Analysis of Port Petrochemical Product Throughput

This paper uses the GRA method to deal with the impact indicators of the throughput of petrochemical products in a port of a city, so as to analyze the correlation between the various impact indicators. After processing by the GRA method, the correlation value of each indicator in the impact indicator system is shown in Fig. 7.

Fig. 7 shows that the grey correlation degree values of the impact indicators on the throughput of petrochemical products in different ports are different. Among these indicators, the grey correlation degree value of x'_{11} indicator is the largest, 0.882. The second is x'_{13}, whose grey correlation value is 0.011 less than x'_{11}. The grey correlation value of x'_{13} is 0.871. The grey correlation degree values of indicators x'_{22}, x'_{12} and x'_{24} are 0.854, 0.849 and 0.817 respectively. The grey correlation degree value of indicator x'_{12} is 0.032 higher than x'_{24}. The index with the lowest grey correlation value of the impact index is x'_{32}, and its grey correlation value is 0.571, which is less than 0.6, that is, the urbanization rate has a weak correlation with the throughput of petrochemical products in a city's port. In addition to the urbanization rate, the other 13 impact indicators have a strong or general correlation with the throughput of petrochemical products in the port of the city. Among them, GDP, total import and export of petrochemical products, coastal berths and the throughput of petrochemical products in the port of a city have a strong correlation. GDP is the most critical impact indicator. Therefore, 13 impact indicators other than urbanization rate are selected as the input data of the prediction model. The training data selects the data from 2006 to 2016. After preprocessing, the ICSO-SVM algorithm is trained through the processed data, and the relevant results are shown in Fig. 8.

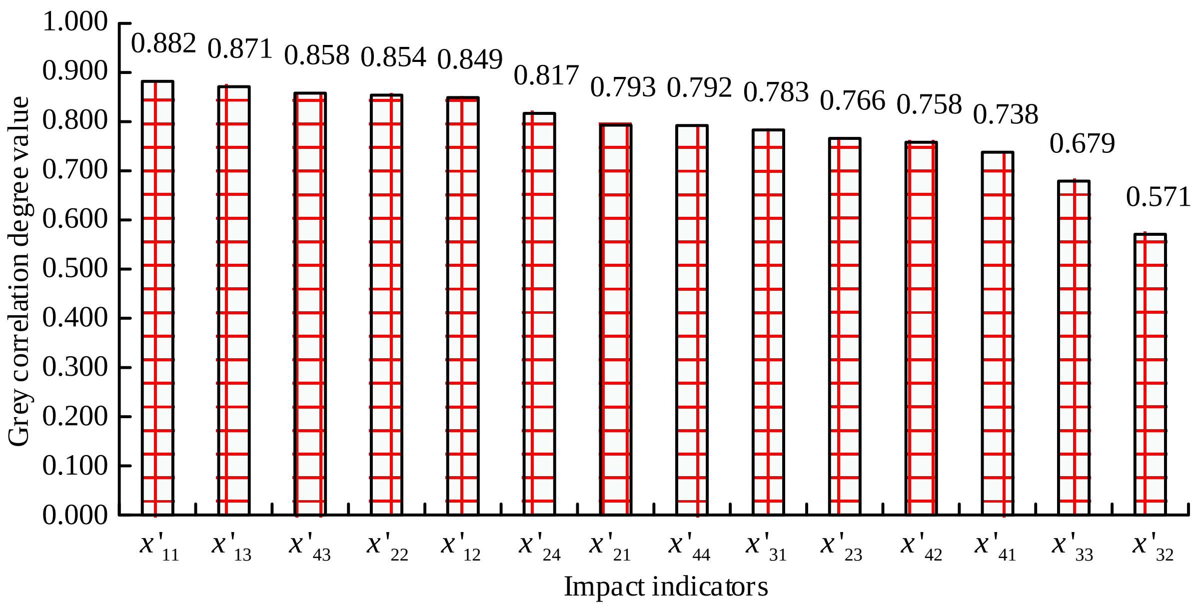
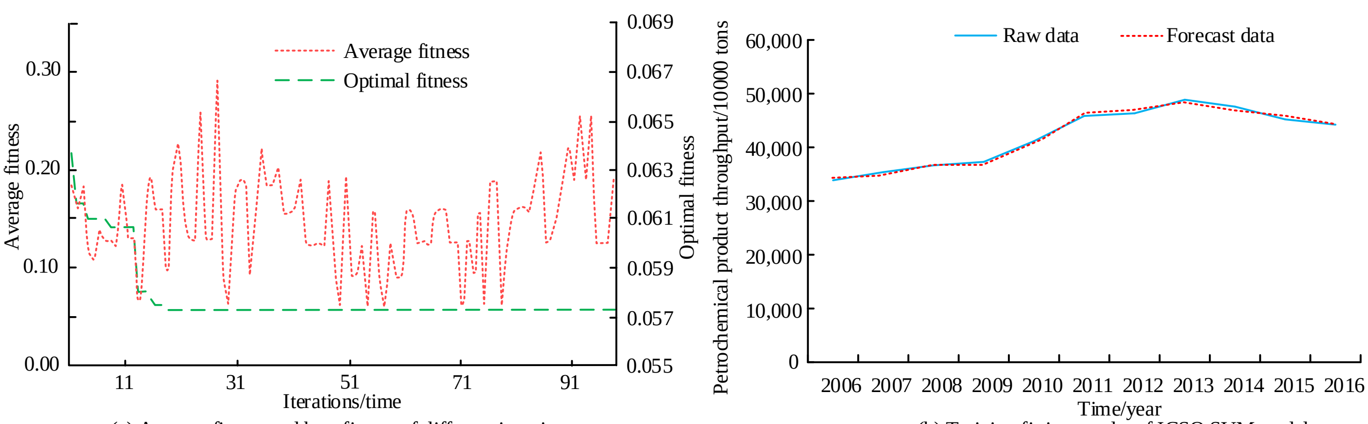


Fig. 7. Grey correlation degree value of impact index.



(a) Average fitness and best fitness of different iterations

(b) Training fitting results of ICSO-SVM model

Fig. 8. Algorithm iterative optimization and training fitting results.

Fig. 8(a) shows the iterative optimization process of ICSO algorithm. Fig. 8(b) shows the training results of ICSO-SVM algorithm. In Fig. 8(a), the average fitness value of the ICSO algorithm is different under different iterations. The optimal fitness decreased slowly and then stabilized. When the iteration number is 12, the average fitness value of the ICSO algorithm is 0.13, which is 0.05 less than the average fitness value of the iteration number of 10. When the iteration number is 28, the average fitness value of ICSO algorithm is the largest, and its average fitness value is 0.24. For the best fitness of ICSO algorithm, when the iteration number is 10, the best fitness value is 0.0607, which is 0.0029 less than that when the iteration number is 1, and the best fitness value of the latter is 0.0636. When the iteration number is 20, the best fitness value of the ICSO algorithm is 0.0572, and the ICSO algorithm begins to converge. Therefore, the optimization speed of the proposed algorithm is fast and the error is low. In Fig. 8(b), according to the line chart of the predicted value of different years and the original data, the two fit well and the fitting result is good. In 2007, the original data was 353.71 million tons, and the forecast data was 348.09 million tons, which was 5.62 million tons less than the former. In 2008, the forecast data was 366.65 million tons, 180000 tons more than the original data. Therefore, the gap between the two is small. The prediction results of different algorithms in the test set are shown in Fig. 9.

In Fig. 9, according to the broken line of the prediction results of different algorithms, the broken line of the prediction results of the ICSO-SVM algorithm is closest to the broken line of the original data. The distance between other polylines and the polyline where the original data is located is different. In 2017, the forecast result of ICSO-SVM algorithm was 478.53 million tons, and the original data was 472.82 million tons, the former was 5.71 million tons more than the latter. In 2019, the prediction results of ICSO-SVM algorithm, CSO-SVM algorithm and PSO-SVM algorithm were 449.73 million tons, 441.03 million tons and 412.6 million tons respectively, while the original data was 453.8 million tons. The gap between the predicted results of ICSO-SVM algorithm and the original data is the smallest. In 2021, the predicted results of SVM algorithm and ICSO-SVM algorithm are 378.56 million tons and 482.05 million tons respectively, which are 110.54 million tons and 7.05 million tons less than the original data. Quantify the prediction error of the algorithm and the relevant results are shown in Fig. 10.

In Fig. 10(a), different algorithms correspond to different MAE and RMSE. In terms of MAE, the MAE of ICSO-SVM algorithm is 762.2, 477.0 smaller than CSO-SVM algorithm, and the MAE of the latter is 1239.2. The MAE of GA-SVM algorithm is 1905.0, and that of SVM algorithm is 2116.4. The maximum MAE of BP algorithm is 2473.0, which is 1710.8 larger than that of ICSO-SVM algorithm, and the latter has the minimum MAE. In terms of RMSE, the minimum RMSE of ICSO-SVM algorithm is 814.7, while the maximum RMSE of BP algorithm is 2792.4. In Fig. 10(b), the minimum MAPE of ICSO-SVM algorithm is 1.05%, which is 0.66% smaller than CSO-SVM algorithm. Therefore, ICSO-SVM algorithm has the best prediction effect. The algorithm without influencing factor analysis is 14-ICSO-SVM algorithm, and the prediction results under different pretreatment methods are shown in Fig. 11.

In Fig. 11(a), the prediction results obtained by the two pretreatment methods differ greatly. Compared with 14-ICSO-SVM algorithm, ICSO-SVM algorithm has smaller MAE and RMSE values. The MAE value of ICSO-SVM algorithm is 762.2, 1016.4 less than that of 14-ICSO-SVM algorithm, and the MAE value of the latter is 1778.6. The RMSE values of ICSO-SVM algorithm and 14-ICSO-SVM algorithm are 814.7 and 1901.7 respectively. In Fig. 11(b), the MAPE values of ICSO-SVM algorithm and 14-ICSO-SVM algorithm are 1.05% and 2.38% respectively, the latter is 1.33% higher than the former. The results further show that the GRA and PCA analysis of the data before the model prediction is conducive to reducing the prediction error. Input the influence factors after pretreatment into the trained model, and predict the throughput of petrochemical products in a city's port from 2023 to 2027 according to different development scenarios. Fig. 12 gives the results.

In Fig. 12, the predicted value under the high-speed development scenario is the largest. Under the same development scenario, the throughput of petrochemical products increased year by year. Under the low-speed development scenario, the forecast result in 2023 is 483.25 million tons, 27.14 million tons less than that in 2025. The forecast result in 2027 is 53.152 million tons. In 2026, the forecast results under the baseline development scenario and the high-speed development scenario are 528.75 million tons and 533.15 million tons respectively, 12.53 million tons and 59.33 million tons less than the corresponding development scenario in 2027.

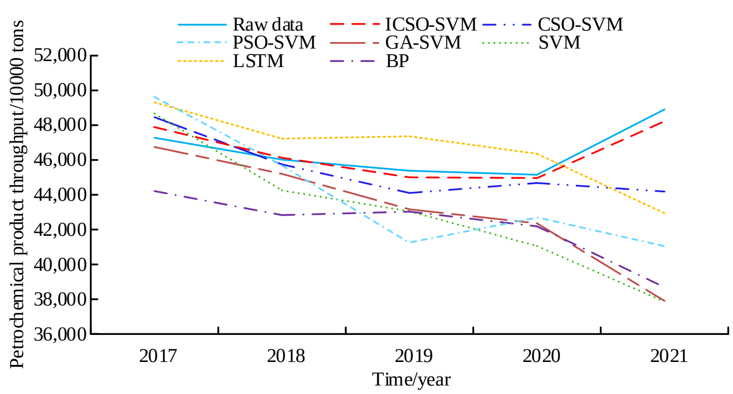
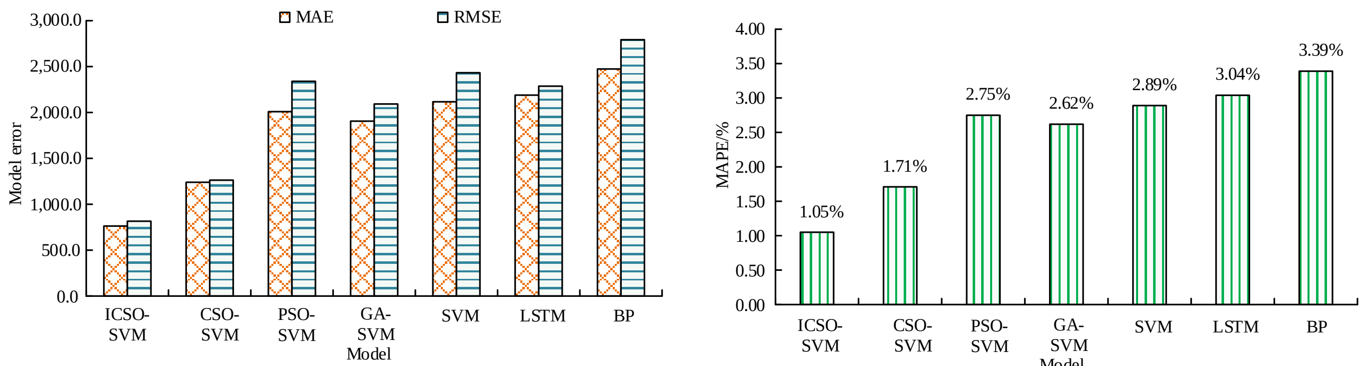


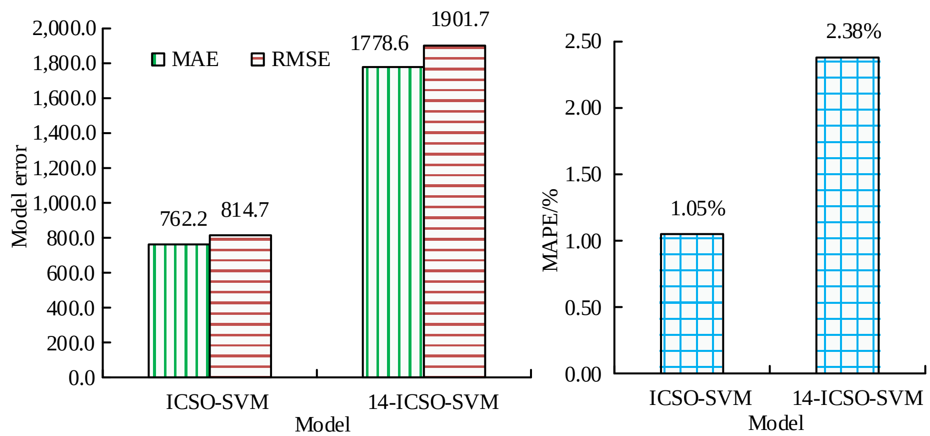
Fig. 9. Prediction results of different algorithms.



(a) MAE and RMSE of different models

(b) MAPE of different models

Fig. 10. Prediction error of different algorithms.



(a) MAE and RMSE of different models

(b) MAPE of different models

Fig. 11. Prediction results under different pretreatment methods.

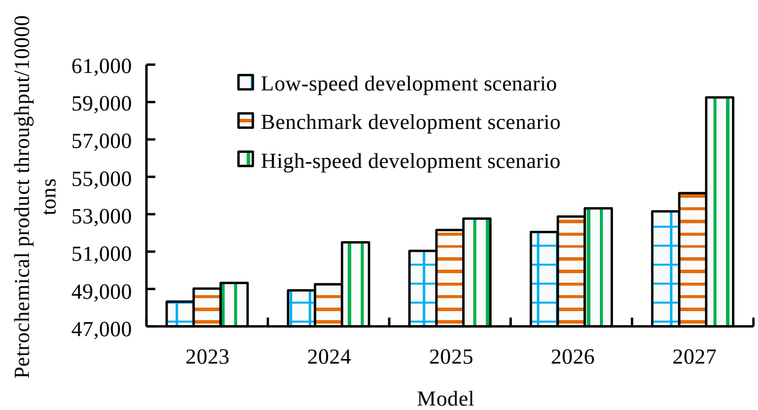


Fig. 12. Prediction results under different scenarios.

In the throughput prediction of petrochemical products at a port in a certain city, the influencing factors of throughput development will affect the prediction results. By screening out key influencing factors, the prediction results are superior to those without considering the influencing factors. Some scholars take into account the influencing factors of FDT when predicting the finish rolling discharge temperature (FDT). From the results, it can be seen that considering the influencing factors, the prediction accuracy of FDT is higher [25]. It can be seen that mining key influencing factors is beneficial for improving prediction accuracy. The model parameters will affect the prediction accuracy of the SVM algorithm, and the more suitable the parameters are, the higher the prediction accuracy of the algorithm. Some scholars choose SVM algorithm to optimize algorithm parameters through simulated annealing algorithm when predicting student grades, in order to obtain the best parameters. After verification, the optimized SVM algorithm has better prediction performance [26]. It can be seen that optimizing parameters can improve the prediction accuracy of the algorithm. When predicting the throughput of petrochemical products, a more comprehensive and scientific consideration of key influencing factors and optimization of SVM algorithm parameters can yield more accurate prediction results, making the obtained results more valuable for reference and providing an effective universal prediction tool for decision-makers.

V. Conclusion

Under the background of the integration of port and industry, to understand the development of the port petrochemical industry, this paper takes the throughput of port petrochemical products as the research object, analyzes its impact factors through the GRA method, and studies its correlation with different impact indicators. The index is sorted and selected based on the grey correlation degree, and the selected influence index is pretreated by PCA method as the input of the prediction model. The research improves the SVM algorithm through the ICSO algorithm, so as to train and test the preprocessed data. At the end of the study, the throughput of petrochemical products in the port of a city in the next five years is predicted. The results show that the optimal fitness value of ICSO algorithm is 0.0572, and the corresponding iteration number is 20. The optimization speed of the algorithm is fast and the error is low. In the training results of ICSO-SVM algorithm, the predicted value is close to the line corresponding to the original data, and the difference between the two data is small. Among the test results of different algorithms, the gap between the predicted results of ICSO-SVM algorithm and the original data is the smallest. In terms of prediction error, ICSO-SVM algorithm has the smallest MAE, RMSE and MAPE. The minimum MAE is 762.2, 1710.8 less than BP algorithm. Its minimum MAPE is 1.06%, which is 0.65% smaller than CSO-SVM algorithm. Among the different preprocessing methods, the prediction result of ICSO-SVM algorithm is better than that of 14-ICSO-SVM algorithm. The former has the lowest MAE, RMSE and MAPE. The MAPE values of ICSO-SVM algorithm and 14-ICSO-SVM algorithm are 1.05% and 2.38% respectively, and the latter is 1.33% higher than the former. Through the prediction of the throughput of petrochemical products in a city's port, its throughput is increasing year by year. Under the high-speed development scenario, the forecast result in 2027 is 592.48 million tons. Thus, the prediction effect of the method used in the article is good. In the future, more scenarios can be set, and the diversity of influencing factors can be considered to make the development scenario prediction more reasonable.

Acknowledgment

The research is supported by: Natural Science Foundation of Liaoning Province “Science and Technology Innovation Ecosystem Construction and Governance Mechanism Research” (2020-YKLH-21); The Key projects of basic scientific research projects in colleges and universities in the Department of Education of Liaoning Province in 2021 “Exploring the path of integrated development of manufacturing and service industries in Liaoning” (LJKR0585); Yingkou Soft Science Plan Project “Research on the Countermeasures of transformation and upgrading of Petrochemical Industry Promoted by ‘Yingkou Port-industry-city’ Integrated development" in 2022 (2022JH2/0100001).

References

[1] Wang Z, Li J, Mu X, et al. A WRF-CMAQ modeling of atmospheric PAH cycling and health risks in the heavy petrochemical industrialized Lanzhou valley, Northwest China. Journal of Cleaner Production, 2021, 291(5):125989.1-125989.9.

[2] Zhang T C, Zheng B Y, Li K, et al. Using CAESAR II software to do stress analysis on centrifugal pump pipeline in Tianjin Port-North China Petrochemical crude oil pipeline project. Xiandai Huagong/modern Chemical Industry, 2017, 37(9):211-212+214.

[3] Maritime, Singapore P. Port Initiatives in Singapore as A Driver of Technology and Growth. Marine Money International, 2018, 34(4):16-19.

[4] Caliskan A, Karaz B. Can market indicators forecast the port throughput?. International Journal of Data Mining Modelling and Management, 2019, 11(1):45-63.

[5] Takamatsu T, Ohtake H, Oozeki T, et al. Regional Solar Irradiance Forecast for Kanto Region by Support Vector Regression Using Forecast of Meso-Ensemble Prediction System. Energies, 2021, 14(11):1-18.

[6] An H, Choi S S, Lee J H. Integrated scheduling of vessel dispatching and port operations in the closed-loop shipping system for transporting petrochemicals. Computers & Chemical Engineering, 2019, 126(JUL.12):485-498.

[7] Zhang N. Research on the Development Strategy of Ningbo Transportation Port & Shipping Industry. Journal of Transportation Technologies, 2019, 09(4):474-488.

[8] Wang Y, Wang N. The role of the port industry in China's national economy: An input–output analysis. Transport Policy, 2019, 78(JUN.):1-7.

[9] Ngoc C T, Xu X, Kim H S, et al. Container port throughput analysis and active management using control theory. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 2022, 236(1):185-195.

[10] Zhang W, Wu Z, Bunn D W. Optimal hybrid framework for carbon price forecasting using time series analysis and least squares support vector machine. Journal of Forecasting, 2022, 41(3):615-632.

[11] Praveena H D, Subhas C, Naidu K R. Classification and discrimination of focal and non-focal EEG signals using hybrid features and support vector machine. International Journal of Advanced Intelligence Paradigms, 2021,18(3):417-437.

[12] Song J, Yu C, Li S. Continuous prediction of onsite PGV for earthquake early warning based on least squares support vector machine (in Chinese). Chinese Journal of Geophysics- Chinese Edition, 2021, 64(2):555-568.

[13] Lai X, Li H, Pan Y. A combined model based on feature selection and support vector machine for PM2.5 prediction. Journal of Intelligent and Fuzzy Systems, 2021, 40(5):10099-10113.

[14] Huang J, Li J, Li M, K Yan. Orthogonal Design-Based Grey Relational Analysis for Influence of Factors on Calcination Temperature in Shaft Calciner. Journal of chemical engineering of Japan, 2019, 52(11):811-821.

[15] Chen C, Wang N, Chen M. Prediction Model of End-point Phosphorus Content in Consteel Electric Furnace Based on PCA-Extra Tree Model. ISIJ International, 2021,6(61):1908-1914.

[16] Luo S, Dai Z, Chen T, H Chen, L Jian A weighted SVM ensemble predictor based on AdaBoost for blast furnace Ironmaking process. Applied Intelligence, 2020, 50(12):1997-2008.

[17] Xu HM, Zhong WJ, Wang CL et al. Quantitative analysis and evaluation of manipulation comfort of tractor gear shifting based on combined methods. Human Factors and Ergonomics in Manufacturing & Service Industries, 2019, 29(4):285-292.

[18] Liu X Y, Zhao Y X, Liu X D, et al. Design of Seamless Knitted Health Care Pants for Knee Arthritis Based on Grey Correlation Analysis. Journal of fiber bioengineering and informatics, 2020,13(2):79-87.

[19] Li X, Zhuo B, Qi X, et al. Grey correlation analysis and path analysis between isoflavones content in Astragali Radix and climate factors. Zhongguo Zhong yao za zhi = Zhongguo zhongyao zazhi = China journal of Chinese materia medica, 2020, 45(14):3407-3413.

[20] Chen Z, Tian K. Optimization of Evaluation Indicators for Driver's

Traffic Literacy: An Improved Principal Component Analysis Method. SAGE Open, 2022, 12(2):242-252.

[21] Yang X, Xiang Y, Jiang B. On multi-fault detection of rolling bearing through probabilistic principal component analysis denoising and Higuchi fractal dimension transformation. Journal of Vibration and Control, 2022, 28(9-10):1214-1226.

[22] Barth J, Katumullage D, Yang C, et al. Classification of Wines Using Principal Component Analysis. Journal of Wine Economics, 2021, 16(1):56-67.

[23] Bhat H F, Wani M A. Novel PSSM-Based Approaches for Gene Identification Using Support Vector Machine. Journal of Information Technology Research, 2021, 14(2):152-173.

[24] Ren J, Zhang B, Zhu X, et al. Damaged cable identification in cable-stayed bridge from bridge deck strain measurements using support vector machine. Advances in Structural Engineering, 2022, 25(4):754-771.

[25] Kang J S, Kim J T, Baek B J. Influence of boundary conditions on FDT prediction and improvement of FDT prediction accuracy in finishing mill process. Asia Life Sciences, 2019(3):1501-1512.

[26] Mahareek E A, Desuky A S, El-Zhni H A. Simulated annealing for svm parameters optimization in student's performance prediction. Bulletin of Electrical Engineering and Informatics, 2021, 10(3): 1211-1219.