Research on gas emission quantity prediction model based on EDA-IGA

Peng Ji [a], Shiliang Shi [a,b], Xingyu Shi [c,*]

[a] College of Resource Environment and Safety Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[b] Provincial Key Laboratory of Safe Mining Techniques of Coal Mines Hunan University of Science and Technology, Taoyuan Road, Yuhu District, Xiangtan, Hunan Province, 411201, China
[c] Changsha University of Science & Technology, Wanjiali Road, Tianxin District, Changsha, Hunan Province, 410114, China

A B S T R A C T

In order to accurately predict the possible gas emission quantity in coal mines, it is proposed to use the multi-thread calculation of the Immune Genetic Algorithm (IGA) and injection of vaccines to improve the accuracy of prediction and combine the Estimation of Distribution Algorithm (EDA) to the distribution probability of excellent populations. Calculating, and selecting excellent populations for iteration, optimize the population generation process of the Immune Genetic Algorithm, so that the population quality is continuously optimized and improved, and the optimal solution is obtained, thereby establishing a gas emission quantity prediction model based on the Immune Genetic Algorithm and Estimation of Distribution Algorithm. Using the 9136 mining face with gas emission hazards in a coal mine from Shandong Province in China as the prediction object, the absolute gas emission quantity is used to scale the gas emission quantity, and it is found that the model can accurately predict the gas emission quantity, which is consistent with the on-site emission unanimous. In the prediction comparison with IGA, it is found that the accuracy of the prediction results has increased by 9.51%, and the number of iterations to achieve the required goal has been reduced by 67%, indicating that the EDA has a better role in optimizing the population update process such as genetic selection of the IGA. Comparing the prediction results of other models, it is found that the prediction accuracy of the EDA-IGA is 94.93%, which is the highest prediction accuracy, indicating that this prediction model can be used as a new method for the prediction of coal mine gas emission. Accurately predicting the gas emission quantity can provide guidance for safe mining in coal mines. The gas emission quantity can also be used as a safety indicator to reduce the possibility of coal mine accidents, ensure the personal safety of coal miners and reduce economic losses in coal mines.

1. Introduction

Coal is the main energy source in China at present, and the demand is constantly increasing. The mining depth of coal mines has also continued to increase, and the disastrous events of gas emission have an increasing trend and are highly harmful, often causing greater casualties and economic losses that are different from other coal mine disasters and bringing huge losses to coal mining. Obstacles and difficulties that need to be addressed. Therefore, in view of the possible gas emission quantity, a prediction method is put

forward, in order to predict the possible gas emission quantity and the possibility of gas emission in advance, to give guidance to the further mining of coal mines, to prevent the occurrence of disasters and accidents, and to ensure the safety of underground mining work, orderly conduct and personnel safety.

At present, there are various methods for predicting the gas emission quantity, and they have achieved good results. Scholars such as Wang Qianyu have studied the application model of gas emission quantity prediction based on support vector machines [1]. Solubility algorithm gas emission quantity prediction model, Cheng Xiaoyu and other scholars have studied the gas emission quantity prediction method of mining face based on random forest and support vector machine [2], but there are still problems that the prediction results are not accurate enough to meet the needs of coal mine safety management. With the development of communication technology and big data technology, mines and ground can realize real-time monitoring and transmission of data, calculation and analysis, and judge the possible gas emission quantity situation underground. Most of the analysis algorithms used, now use intelligent algorithms such as machine learning and deep learning and have achieved some beneficial effects. As an immune algorithm, the Immune Genetic Algorithm (IGA)is because it imitates the biological immune process and the immune antibody [3]. The characteristics of memory, which are consistent with the characteristics of when the gas emission quantity and the size of the gas emission quantity are uncertain, and its steps are simple, stable, and have the characteristics of learning and multi-threaded calculations, which can be used for gas monitoring [4]. Because of the imitation of the immune process, this method is consistent with the uncertainty of gas emission, so it can quickly capture the change of gas emission, compared with traditional prediction models such as particle swarm algorithm, BP neural network, etc. The efficiency is higher, and the characteristics of the algorithm itself are more suitable for the needs. The prediction of the gas emission quantity, through the methods of mutation and crossover and vaccination, retains excellent populations, speeds up the convergence of the algorithm, and saves calculation time, but it is difficult to determine the direction of convergence, vaccine injection alone, it is difficult to guarantee the memory capacity of the algorithm [5]. Therefore, the Estimation of Distribution Algorithm (EDA)is introduced into the IGA, by establishing a probability model, analyzing and comparing excellent groups, constructing a distribution probability model of the gas emission quantity prediction characteristic variable [6], and generating new populations, and through repeated iterations, the quality of the population is continuously improved. Optimizing updates, thereby controlling the convergence direction and memory properties of the algorithm. Under the optimization of the EDA, through the method of probability distribution, the high-affinity antibodies produced by IGA can be quickly retained and distributed, occupying the space of low-affinity antibodies. Compared with the single use of IGA, the model runs faster speed up, and the direction of convergence is controlled.

Based on this, a gas emission quantity prediction model based on the Immune Genetic Algorithm and the Estimation of Distribution Algorithm was established, and a working face with gas emission risk in a coal mine from Shandong Province in China was used as the experimental object to verify the prediction accuracy of the model. It is found that the model has higher prediction accuracy, which shows that the model can be used as an effective prediction method for the prediction of gas emission quantity in coal mines and can be used to guide the engineering practice of coal mines.

2. Algorithm selection

2.1. Genetic algorithm

The Genetic Algorithm (GA) was born in the 1970s [7]. It is an intelligent search method that emerged by drawing on the natural selection of biological nature and the law of genetic biological evolution [8]. Its genetic characteristics are in line with the memory
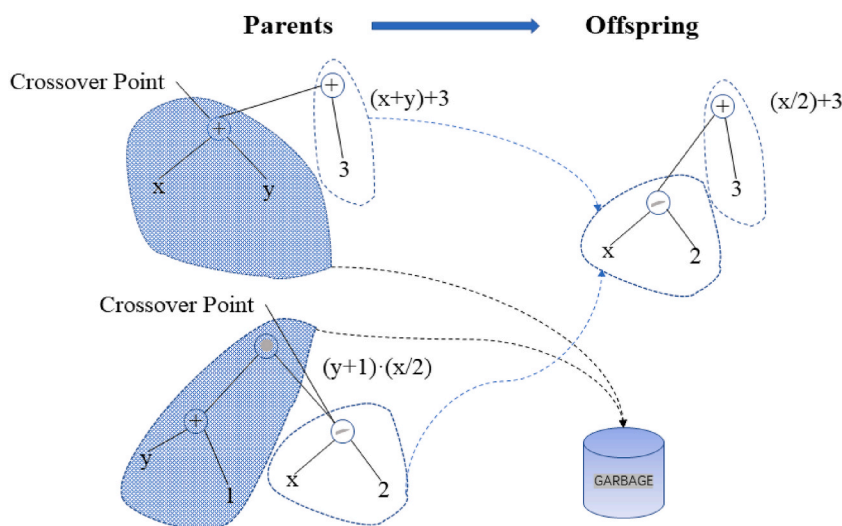

Fig. 1. Principle of crossover operator.

requirements of the time series characteristics of gas emission data, and the characteristics of gas emission in the next period can be inferred from the characteristics of gas emission quantity in the previous period [9]. The algorithm generates a random initial solution, and through repeated selection, crossover, and mutation, the global optimal solution is obtained. The solution process of the genetic algorithm is a simulation of the natural survival of the fittest principle. According to the degree of fitness, select and retain excellent individuals with high fitness, and discard and eliminate individuals with low fitness. This selection method is the way of roulette [10]. It is widely used in the optimal solution of planning, therefore it is convenient for fast solutions, and then estimates the possibility of the existence of the object in the next time.

The crossover and mutation performed by the algorithm are carried out according to the following rules. In the process of crossover, the partial genes of the two chromosomes select two individuals as parents in the previous generation population according to the selection probability equal to the fitness, randomly select their tree structure nodes, and exchange the two children tree with this node as the root node, thus forming two new individuals [11]. The proportion of the number of individuals formed in this way to the size of the group is Pc, and the crossover process is shown in Fig. 1.

There are only three crossover methods in Table 1. During the crossover process, a pair of chromosomes are randomly selected, and one or several positions of the string are randomly selected for crossover according to the length of the string [12]. In this paper, the method of multi-position crossover is selected. The crossover operator.

After the crossover operator runs, the genetic algorithm will carry out the mutation operation, as shown in Fig. 2, selecting an individual in the previous generation group as the parent according to the selection probability proportional to the fitness, randomly select its tree structure node, and randomly generate another tree structure [13]. The function body replaces the subtree with this node as the root node, thereby generating a new individual. The proportion of the number of individuals thus generated to the population size is P_m.

As shown in Table 2, the mutation operators in the mutation operation can be divided into basic bit mutation, uniform mutation, boundary mutation, non-uniform mutation, and so on.

In addition to the crossover operator and mutation operator, there is also the use of an elite operator and replication operator in the algorithm. The elite operator is several individuals or a certain proportion of individuals with the highest fitness in the previous generation group as new individuals directly put into the next generation group, the proportion of the number of individuals in this part is Pe, and the elite operator can effectively avoid the loss of the optimal individual that may occur due to the inclusion of the genetic algorithm. It has part of the function of memory. The replication operator selects individuals from the previous generation group into the next generation group according to a certain proportion of the selection probability of fitness, and the proportion of the number of individuals thus generated to the size of the group is Pr [14]. The addition of the two algorithms keeps the diversity of antibodies while avoiding the premature problem of the algorithm. Before the algorithm operation, it is necessary to set the population proportion that the operator needs to operate, and the sum of the proportions is Pc+ Pm+ Pe + Pr = 1.

2.2. Immune Genetic Algorithm

Due to the simple steps, strong stability, and functional characteristics of learning and multi-threaded computing, the genetic algorithm has been widely and fully applied in the engineering field. How to make the algorithm effectively transfer the properties of the previous generation to the next generation and solve the problem of the convergence characteristics of the genetic algorithm, scholars have established the Immune Genetic Algorithm (IGA) [15]. Memory and convergence problems are addressed by vaccinating. In IGA, each individual in the population is regarded as an antibody, and the antigen is regarded as a problem to be solved. At the same time, the fitness in the genetic algorithm is regarded as the affinity of the antigen, and the genotype of the individual is regarded as the antibody's affinity. Antibody determinants, also known as antibody strings. The calculation flow chart of IGA is shown in Fig. 3, and its specific steps are as follows.

(1) Creating an initial random population M_1;
(2) Drawing vaccines based on previous experience and knowledge;
(3) If the optimal individual appears in the current k-th generation group M_k, the algorithm stops, and if it does not appear, the algorithm continues;
(4) In the algorithm, the crossover operation is performed on the group M_k to form a temporary group N_k;
(5) The mutation operation of group N_k within the algorithm forms a new temporary group V_k;
(6) Produce W_k by injecting vaccines against V_k;

Table 1
Example of crossover operation.

Chromosome	Crossway	Cross result
11010010	Multipoint crossover (position 2, 4, 6)	10000110
10011010		11001110
10010101	Two-point cross (position 3, 7)	10110100
01001110		01101100
00100110	Single Point Crossing (position 5)	00101110
11011010		11010010
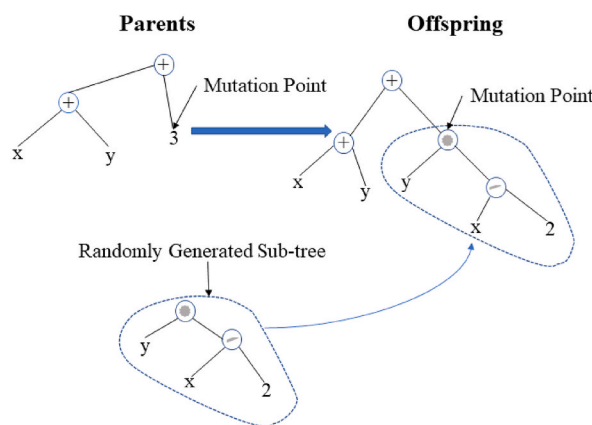
Fig. 2. Principle of the mutation operator.

Table 2
Example of mutation operation.

Chromosome	Crossway	Cross result
11010110	Mutation at position 2	10010110
11010110	Mutation at position 4	11000110
11010110	Mutation at positions 3and5	11111110

(7) Perform immune selection on W_k to generate a mature next-generation population M_k, and then enter step (3);

The vaccine in the algorithm is a variant obtained by randomly mutating several bits of each individual, which is obtained according to prior knowledge and has a higher probability to obtain a higher fitness [16]. The so-called vaccine injection means that the characters in all positions of the mutated individual in this algorithm are different from the optimal solution, then the probability of mutating into this new individual is 0. If the characters in all positions of the mutated individual are the same as the optimal solution, then the probability of mutating into this new individual is 1. Then, there will be $n_\alpha (n_i = n*\alpha)$ antibodies in the population A = {a_1, a_2, …,a_n}that are selected to be vaccinated, where $\alpha \in (0,1)$. After the vaccine injection is over, during the running process of the algorithm, if the fitness of the new individual is not as good as that of the old individual, the old individual will continue to be used for subsequent operations, and the annealing selection will be used in the current descendant population, using equation (1) roulette the idea of selecting an individual a_i (i = 1, …,n) into the new parent population M_{k+1} is the immune selection of the algorithm.

$$P(a_i) = \frac{e^{f(a_i)/U_k}}{\sum_{i=1}^{n} e^{f(a_i)/U_k}}$$

(1)

In Eq. (1), $f_{(ai)}$ is the fitness value of individual ai, and U_k is the temperature variable that gradually tends to 0.

The IGA is generated by combining a genetic algorithm and an immune algorithm. The basic flow of the algorithm is shown in Fig. 3. The algorithm of vaccine injection is added based on a genetic algorithm. It not only has the advantages of parallel computing and self-adaptation of the genetic algorithm but also has the characteristics of learning memory and self-identification and non-self identification of the immune algorithm [17]. The writing of the vaccine injection algorithm makes the algorithm a memory function, which is convenient for immune selection and suppression and makes the algorithm converge in a controllable direction. Given the above advantages of the Immune Genetic Algorithm, it is used to predict the gas emission quantity.

3. Estimation of Distribution Algorithms

To meet the demand for accuracy in gas emission quantity prediction, the Immune Genetic Algorithm is optimized using the Estimation of Distribution Algorithms (EDA) [18]. The Estimation of Distribution Algorithms is a population-based stochastic optimization algorithm and a probability based genetic algorithm, which is consistent with the data decomposition method in this paper [19]. Establishing a probability model and sampling in the algorithm are the core steps, and it is also the biggest difference from the genetic algorithm. In the algorithm, the mutation and immune operation of the genetic algorithm are replaced by analyzing the variables of the comparatively excellent population, and the distribution probability model of these variables is constructed, thereby generating a new population, so the quality of the new population is also better than the original population, and the population undergoes repeated iterations [20]. The quality is continuously optimized and updated, realizing search optimization similar to scheduling.
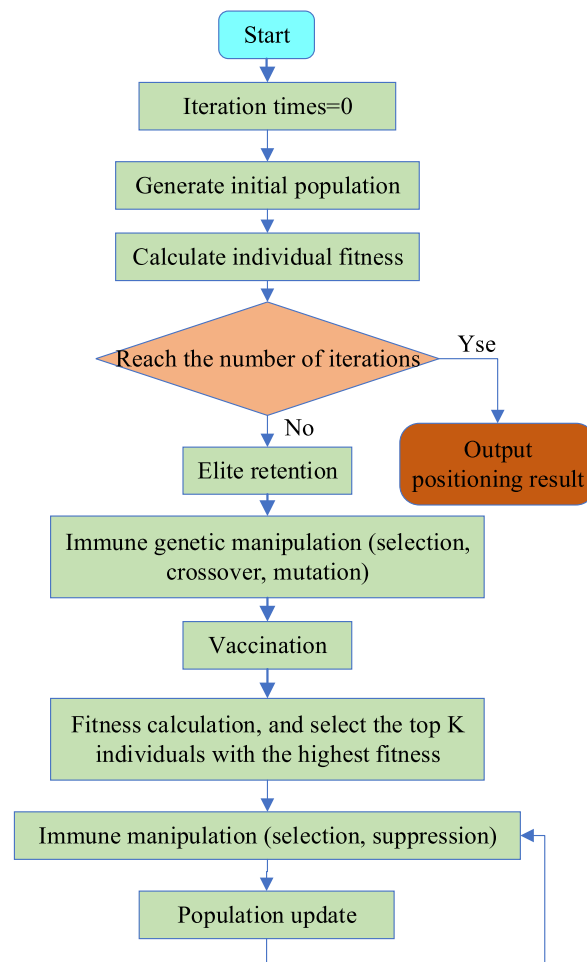
Fig. 3. IGA flow chart.

In Fig. 4, (a)Population-based methods, (b)Evolutionary Strategies, (c)EDAs. The green circle is the current population generation and sampling area, and the blue dots are the samples with good evaluation. The orange-yellow dots are the poorly rated samples [21]. The green circles are the newly generated generation and the sampling area, the blue dots are the newly generated samples, and the
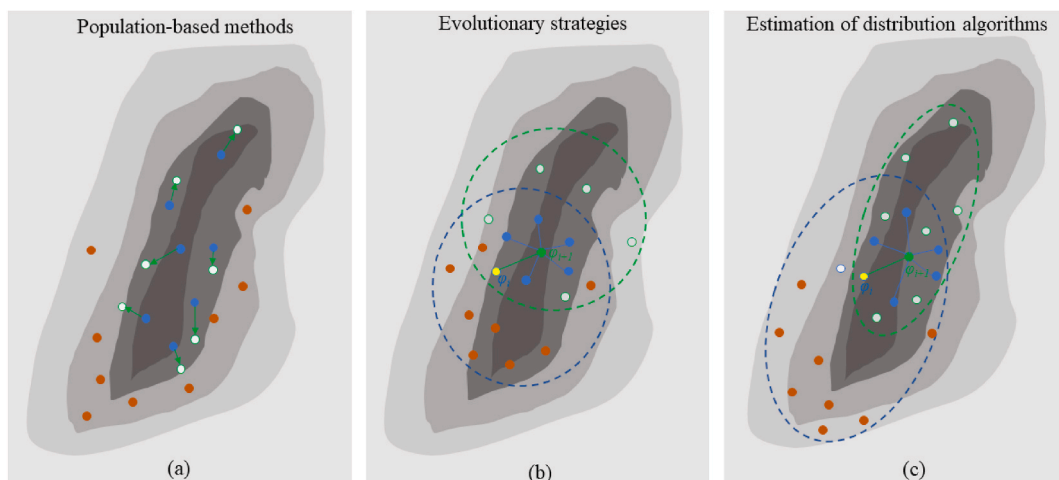

Fig. 4. One iteration of evolutionary methods.

green hollow dots are the samples that have not yet been evaluated, yellow is also not yet evaluated [22]. In ES, it is obtained from an optimum guess and sampling from fixed Gaussian noise. In EDAs, Gaussian noise is tuned using Covariance Matrix Adaptation [23].

With the help of this advantage of EDA, optimize the search function of the Immune Genetic Algorithm, improve its calculation speed and memory optimization ability, and establish a hybrid method based on EDA-IGA to predict the possible gas emission [24]. As shown in Fig. 5, in EDA-IGA, new populations are generated jointly by IGA and EDA. And IGA and EDA share an original population to generate two different sub-populations according to their respective algorithm mechanisms, and then mix these two sub-populations to complete a generation of evolution [25].

4. Establishment of prediction model based on EDA-IGA

On the basis of the introduction of the IGA algorithm and the EDA algorithm above, the data collected from the coal mine underground monitors are normalized, and the gas emission quantity is predicted by EDA-IGA. The basic idea is to use EDA to optimize the population generation process of IGA and design the evaluation function parameters of the population generation solution according to the actual situation of gas emission in the previous period, so as to improve the performance of the algorithm.

4.1. Coupling relationship between demand and model

In order to make the model more explicit how the gas emission quantity is predicted step-by-step, and how the real-world data operates within the algorithm as a biological component. In Table 3, we have coupled the demand for gas emission quantity prediction with the operating mechanism in the model and the operating points in it, so that the corresponding relationship between the forecast demand and the establishment of the model is clear and clear. In the table, the biological immune process in IGA is clearly compared with the prediction process of gas emission quantity, so that the model can fulfill all requirements.

4.2. Data processing

In gas emission quantity prediction, since the selected data is time series, different stages show different rules, showing the same fluctuation characteristics as seasons, therefore, time series data can usually be decomposed into three parts: trend items, periodic items, and irregular residual items [26]. The daily outburst of the working face, the daily advance rate, the gas content of the coal seam and surrounding rock, the surface atmospheric pressure, the geological structure, the mining depth, and the air volume will all have an impact on the gas emission. Gas emission has a trend effect on daily outburst, daily advance and air volume have periodic effects on the gas emission quantity of the working face, geological structure changes, atmospheric pressure changes, periodic pressure, etc. The influence of mutations, the influence of these factors can be characterized in the time series data of gas emission quantity. Direct use of raw data will lead to insufficient utilization of data. Therefore, the time series data decomposition method Seasonal and Trend decomposition using Loess (STL) using locally weighted regression for periodic trend decomposition is used to decompose time series data of gas emission [27]. Since the overall timing diagram of the analyzed gas emission data is in a relatively stable state, the additive STL time series is selected to decompose the data [28]. Each part of the decomposition is shown in Eq. (2).
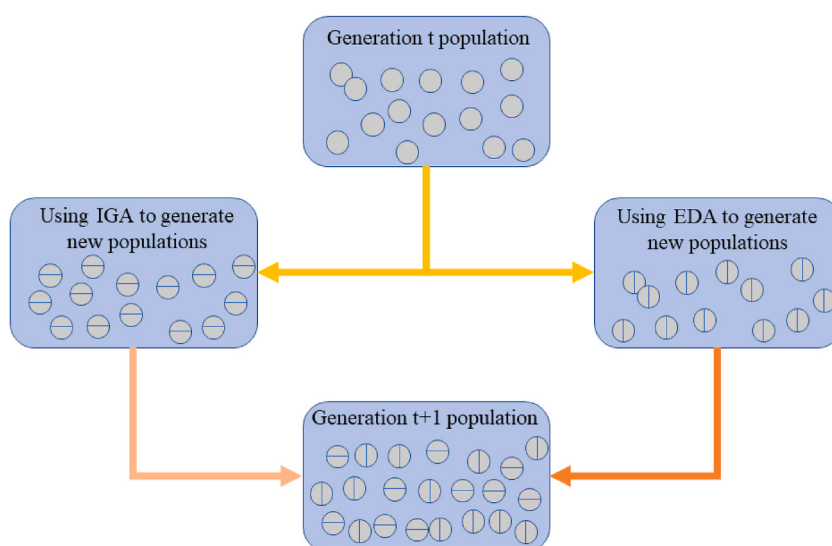
$$Y_t = T_t + S_t + R_t \tag{2}$$

Fig. 5. EDA-IGA hybrid mechanism.

Table 3
Comparison of biological immune process and algorithm model.

Inside the IGA	Gas emission quantity prediction
Antigen/self	Gas emission quantity index factor training set data
Antibody/noself	Gas emission quantity index factor prediction set data
T cell	Detector
Foreign proteins	Predicting the concentration of supercritical abnormal data
Antigen identification	Abnormal component identification and discovery
Cell clone	Antibodies with high clonal detection rates

In Eq. (2), T_t represents the trend value at time t, S_t represents the value of the periodic term at time t, and R_t represents the value of the irregular residual term at time t. STL key parameter setting: t. window to control the change speed of the trend effect is set to 13, s. window to control the change speed of the cycling effect is set to "period", and the Loess process uses robust fitting, that is, robust = T.

In the actual model training and prediction process, since the obtained irregular residual items data with large fluctuations and more peaks is not conducive to model training and learning, after data processing, the third type of irregular residual item R_t individual data are excluded. The flow chart of the established gas emission quantity prediction model based on EDA-IGA is shown in Fig. 6.

The specific operation steps of the model are as follows.

(1) Collect gas emission quantity related index data from coal mines, and homogenize the data so that the model can be read. The model reads the data, and implements mean interpolation measures for missing and other inconsistent data.
(2) Using the STL method to decompose the data into four parts, Irregular residual items, trend items and periodic items, data that does not match the data set are discarded.
(3) Input the data into the EDA-IGA model for prediction, and the decomposed data form the training set and the prediction set respectively, Initialize the IGA parameter and population. Affinity calculation, perform immunization processes against high-affinity antibodies, such as selection, crossover, mutation, etc.
(4) Then, the EDA model selects high-affinity antibodies, performs a probability distribution process, and randomly distributes high-affinity antibodies. Subsequently, the antibodies of the two models were mixed.
(5) Superposition component the EDA and IGA prediction results, get the final gas emission quantity prediction result.

4.3. Diversity evaluation of model solutions

In order to make the prediction results accurate, the model does not deviate in a single direction, and the diversity of the population is evaluated. The measures are as follows: The diversity evaluation of the antibody is carried out by Eq. (3), and the expected
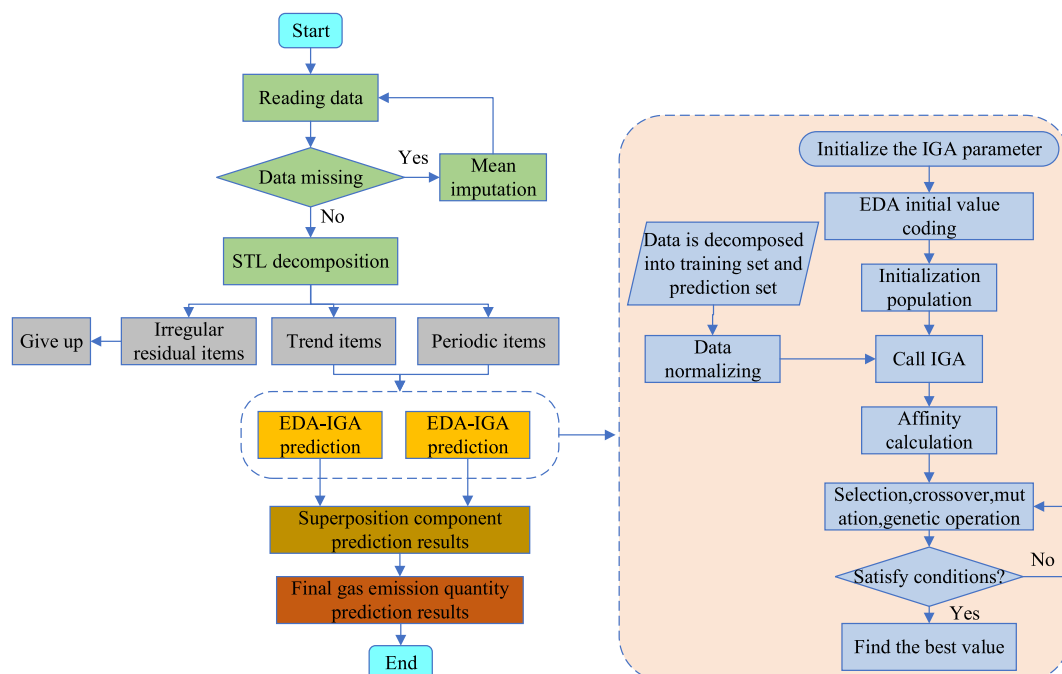

Fig. 6. Flow chart of gas emission quantity prediction model.

reproduction rate R of the antibody is jointly determined by the individual fitness $f_{(ai)}$ and the antibody concentration T. Based on this, the algorithm suppresses antibodies with high concentrations and low fitness, and promotes antibodies with low concentrations and high fitness, and evaluates the diversity of the population at one time.

The expression of antibody concentration T is expressed by the proportion of similar antibodies in the population with the total number of antibodies Q.

$$T = \sum_{n \in Q} M/Q \tag{3}$$

Evaluating the diversity of antibodies in the model to avoid the convergence direction deviation caused by a single antibody, so as to ensure that the prediction results of the model are sufficiently accurate.

5. Verify the model

5.1. Data collection and processing

In this paper, the 9136 mining face of a mine with gas emission from Shandong Province in China is used as the prediction object and the data are collected. In this working face, the average thickness of the coal seam is 2.8 m, the average inclination angle of the coal seam is 2.1°, and the gas content is 0.53–6.73 m³/t. The collected gas emission data of the working face is a total of 400 sets of data from June 2019 to July 2020, as shown in Table 4, including gas air displacement, gas drainage quantity, and absolute gas emission quantity data.

Using STL to split the data, divide the data into trend items, periodic items and irregular residual items, remove the data in the irregular residual items, and remove the irregular residual data for the missing data in the data set. The method of post-mean imputation. Randomly split the selected data into two different training prediction sets. For the first group, select 180 groups as the training set and 220 groups as the prediction set; for the second group, select 100 groups as the training set and 300 groups as the prediction set. Training the model.

The kernel function of Gaussian probability distribution is used in EDA to optimize the hyperparameters of each component EDA-IGA model, and the larger the kernel function Gamma, the narrower the Gaussian distribution. The smaller the Gamma, the wider the Gaussian distribution. Gamma is equivalent to adjusting the complexity of the model. The smaller the Gamma value, the lower the model complexity. The higher the Gamma value, the greater the model complexity. c is the penalty coefficient, that is the tolerance for sample misclassification. The higher c is, the more intolerable it is to be wrong, and it is easy to overfit. The smaller c is, the easier it is to underfit. Optimizing through training, and obtaining the optimal value is shown in Table 5. Determine the best model parameters for each component and complete the construction of the gas emission prediction model based on EDA-IGA. It is worth noting that for different problems, different model parameters need to be set according to the characteristic requirements of the problem and the characteristics of the data in order to obtain the optimal results.

In order to verify the prediction effect of the built model, the forecast set data is used to predict the absolute gas emission quantity of the mining face in the future time period, and the first set of data is used as the prediction object to predict the gas emission quantity in the next two months 60 days. The prediction results of the two components of the trend item and the period item and the pairing error generated by the prediction are shown in Fig. 7, and the absolute error value is within the allowable range of the required accuracy of the prediction.

The prediction results of the two components are superimposed and reconstructed to obtain the required absolute gas emission quantity prediction value, as shown in Table 6. It can be obtained from the data in the table that the average absolute error between the observed value and the predicted value is 0.61, and the error accounts for 0.010 of the average value of the measured value, indicating that the predicted value is almost consistent with the measured value, and it is considered that the prediction of the model is effective, and the accuracy can be accept.

And the curve shown in Fig. 8, the final gas emission quantity prediction value curve approximately coincides with the actual gas emission quantity curve, and the curve fitting degree is high. The absolute error range of the prediction results is 0.0591–1.83 m³/min, and the absolute error is 0.141%–1.794%. The error meets the gas emission quantity prediction accuracy requirements. The predicted two months gas emission quantity trend of the mining face is consistent with the actual situation is consistent, indicating that the established gas emission quantity prediction model based on EDA-IGA is effective.

Table 4
Raw data of gas emission in 9136 working face.

b	Gas air displacement/(m³/min)	Gas drainage quantity/(m³/min)	Absolute gas emission quantity/(m³/min)
1	6.48	30.01	38.21
2	6.01	30.82	37.96
3	7.32	30.93	38.43
4	7.11	32.11	42.32
5	4.83	31.03	45.34
⋮	⋮	⋮	⋮
99	5.35	39.98	47.53
400	5.47	40.32	46.87

Table 5
The hyperparameter optimization value of each component in the model.

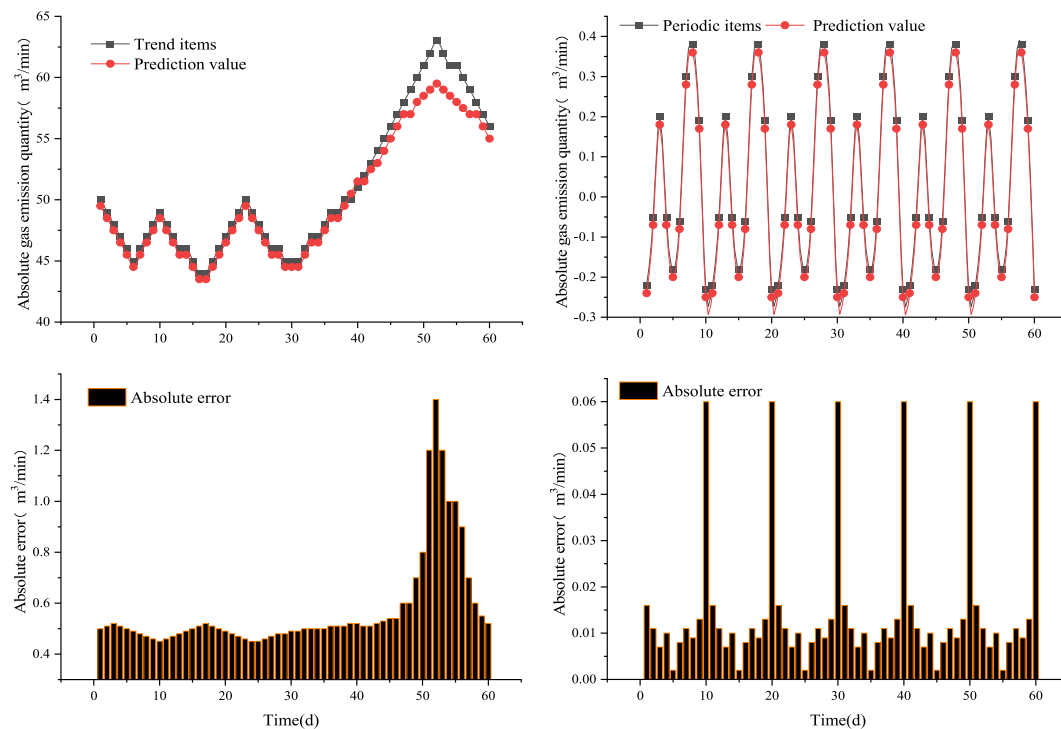| Component	The first group of optimization values (best)		The second group of optimization values (best)	
	Penalty coefficient c	Kernel function Gamma	Penalty coefficient c	Kernel function Gamma
Trend item	6.73	0.08	6.52	0.03
Periodic term	6.32	0.61	1.91	2.23

Fig. 7. Prediction results of time series decomposition component model.

Table 6
Comparison table of predicted value and observed value.

b	EDA-IGA final prediction value	Observed value	Absolute error
1	51.5	52	0.5
2	49.5	50	0.51
3	48.5	49	0.52
4	47.5	48	0.51
⋮	⋮	⋮	⋮
59	58.5	59	0.55
60	55.5	56	0.52

5.2. Performance comparison of different algorithms

In this paper, the Quantum Immune Genetic Algorithm (QIGA), IGA and EDA-IGA are used to predict the gas emission quantity, and the convergence speed and optimization ability of the three algorithms are compared and analyzed, and the performance of the EDA-IGA algorithm is tested. The population size of the three algorithms is set to 40, and the maximum number of iterations is 80. The memory capacity of IGA is set to 10, each parameter of the solution is represented by 20 binary bits, the expected reproduction rate is 0.9, and the similarity coefficient is 0.75. The setting parameters of QIGA and IGA are the same as those of EDA-IGA. The actual comparison is shown in Fig. 9. It can be seen that the three algorithms can accurately predict the gas emission quantity. It can be seen intuitively from the figure that the predicted result curve of the EDA-IGA model is consistent with the measured value curve, and the QIGA and IGA models have different degrees of deviation in the prediction results, and the prediction results are inaccurate.

Therefore, the prediction performance of the three models is compared in terms of the average number of iterations, the number of successful optimizations, and the average search time, as well as the Mean Absolute Error (MAE), the Mean Absolute Percentage Error
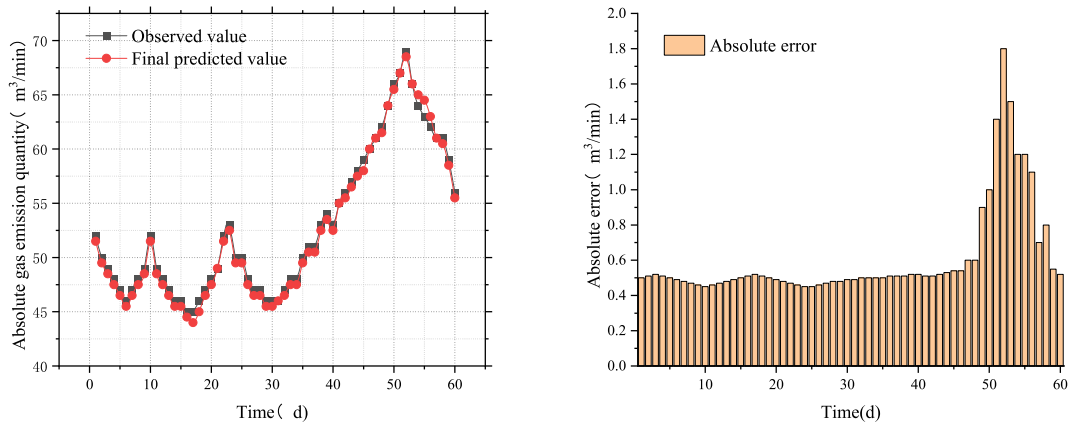
Fig. 8. Comparison of model prediction results and measured value curves.

(MAPE), and the Root Mean Square Error (RMSE). Its calculation equations are (4)–(6):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \widehat{y}_i - y_z \right| \tag{4}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\widehat{y}_i - y_i}{y_i} \right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{y}_i - y_i \right)^2} \tag{6}$$

For the performance comparison of the three algorithms, the population size of the three algorithms is set to 40, the maximum number of iterations is 80, and other parameters remain unchanged. Comparing the performance data of each prediction model in Table 7, it is found that EDA-IGA has the highest prediction accuracy rate of 94.93%, which is 9.51% higher than IGA and 4.61% higher than QIGA. And the number of iterations is reduced by 67% compared with IGA. However, the average search time is 64.5s longer than that of IGA, which is caused by the superposition of population updates of the two algorithms IGA and EDA, but it has the highest number of successful optimizations. It shows that EDA has good optimization performance on the genetic and mutation selection process of IGA population update, and it is a good performance prediction method for coal mine gas emission quantity.

6. Conclusions and prospects

(1) The EDA-IGA model for the prediction of gas emission in coal mining face is proposed, using STL to decompose the absolute gas emission data into trend items, periodic items and irregular fluctuation items, making full use of the data characteristics, through EDA parameters optimized IGA establishes a prediction model, which reduces data complexity, optimizes stationarity, improves prediction accuracy, and the prediction results are consistent with the actual gas emission quantity.

(2) In the prediction comparison with IGA, it is found that the accuracy of the prediction results is 9.51% higher, and the number of iterations to achieve the required goal is reduced by 67%, which shows that EDA has a better optimization effect on the population update process such as genetic selection of IGA.

(3) Comparing the prediction results of the three models of QIGA, IGA and EDA-IGA, compared with the measured values, it is found that the prediction accuracy of EDA-IGA is 94.93%, which is higher than that of QIGA is 90.32% and IGA is 85.42% are all high, indicating that EDA-IGA is a better prediction model for coal mine gas emission.

(4) Since the algorithm needs multiple attempts to find the optimal parameters in the early stage, it takes a lot of time, which is the shortcoming of the model, but once the rightmost parameter is found, it will be done once and for all. In the further research, we will explore more data processing methods and prediction methods suitable for gas emission quantity, in order to find a more convenient and effective prediction methods.

Author contribution statement

Peng Ji: conceived and designed the experiments; performed the experiments; analyzed and interpreted the data; contributed reagents, materials, analysis tools or data; wrote the paper.

Shiliang Shi, Xingyu Shi: conceived and designed the experiments; contributed reagents, materials, analysis tools or data; wrote the paper.
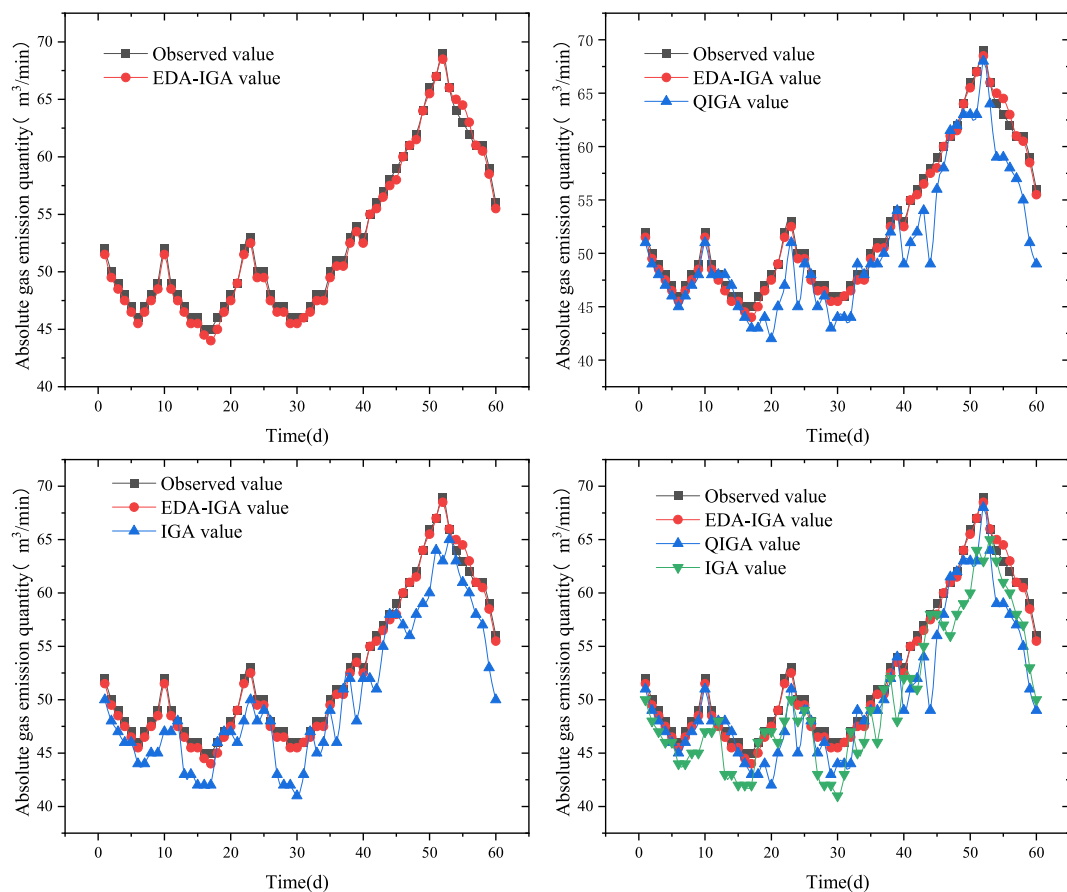
Fig. 9. Comparison curves of predicted values of three algorithms.

Table 7
Comparison of prediction effects of different models.

Model	EDA-IGA	IGA	QIGA
Mean number of iterations	28	85	46
Number of search successes	33	25	32
Mean search time/s	732.9	668.4	753.9
MAE	1.58	2.09	2.32
MAPE	2.93	3.88	4.35
RMSE	1.87	2.63	2.33
Prediction accuracy/%	94.93	85.42	90.32

Data availability statement

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] Wang,Q.,Zhang, Y., Gas emission analysis and prediction of mining working face based on PLSMA-SVR model.J. Liaoning Inst. Technol.. Vol.29 No.6.

[2] X. Cheng, A. Zhou, Y. Guo, C. Cheng, D. Li, Prediction method of gas emission based on random forest and support vector machine, Saf. Coal Mine 53 (10) (2022).

[3] F. Chen, B. Tang, R. Chen, A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, Measurement 46 (2013) 220–232, https://doi.org/10.1016/j.measurement.2012.06.009.

[4] H. Han, Y.-S. Ding, K.-R. Hao, X. Liang, An evolutionary particle filter with the immune genetic algorithm for intelligent video target tracking, Comput. Math. Appl. 62 (2011) 2685–2695, https://doi.org/10.1016/j.camwa.2011.06.050.

[5] Z. Zhou, D. Luo, J. Shao, Y. Xu, Y. You, Immune genetic algorithm based multi-UAV cooperative target search with event-triggered mechanism, Phys. Commun. 41 (2020), 101103, https://doi.org/10.1016/j.phycom.2020.101103.

[6] Y. Du, J. Li, C. Luo, L. Meng, A hybrid estimation of distribution algorithm for distributed flexible job shop scheduling with crane transportations, Swarm Evol. Comput. 62 (2021), 100861, https://doi.org/10.1016/j.swevo.2021.100861.

[7] B. Yang, C. Zeng, D. Li, Z. Guo, Y. Chen, H. Shu, P. Cao, Z. Li, Improved immune genetic algorithm based TEG system reconfiguration under non-uniform temperature distribution, Appl. Energy 325 (2022), 119691, https://doi.org/10.1016/j.apenergy.2022.119691.

[8] G.-C. Liao, Integrated isolation niche and immune genetic algorithm for solving bid-based dynamic economic dispatch, Int. J. Electr. Power Energy Syst. 42 (2012) 264–275, https://doi.org/10.1016/j.ijepes.2012.03.005.

[9] L. Tao, X. Kong, W. Zhong, F. Qian, Modified self-adaptive immune genetic algorithm for optimization of combustion side reaction of p-xylene oxidation, Chin. J. Chem. Eng. 20 (2012) 1047–1052, https://doi.org/10.1016/S1004-9541(12)60585-0.

[10] Z. Zhang, X. Wang, J. Lu, Multi-objective immune genetic algorithm solving nonlinear interval-valued programming, Eng. Appl. Artif. Intell. 67 (2018) 235–245, https://doi.org/10.1016/j.engappai.2017.10.004.

[11] Md Asadujjaman, H.F. Rahman, R.K. Chakrabortty, M.J. Ryan, Multi-operator immune genetic algorithm for project scheduling with discounted cash flows, Expert Syst. Appl. 195 (2022), 116589, https://doi.org/10.1016/j.eswa.2022.116589.

[12] H. Lin, L. Sun, Searching globally optimal parameter sequence for defeating Runge phenomenon by immunity genetic algorithm, Appl. Math. Comput. 264 (2015) 85–98, https://doi.org/10.1016/j.amc.2015.04.069.

[13] R.J. Kuo, Y.H. Lee, F.E. Zulvia, F.C. Tien, Solving bi-level linear programming problem through hybrid of immune genetic algorithm and particle swarm optimization algorithm, Appl. Math. Comput. 266 (2015) 1013–1026, https://doi.org/10.1016/j.amc.2015.06.025.

[14] M. Duma, B. Twala, Sparseness reduction in collaborative filtering using a nearest neighbour artificial immune system with genetic algorithms, Expert Syst. Appl. 132 (2019) 110–125, https://doi.org/10.1016/j.eswa.2019.04.034.

[15] H.Y. Guo, Z.L. Li, Structural damage identification based on Bayesian theory and improved immune genetic algorithm, Expert Syst. Appl. 39 (2012) 6426–6434, https://doi.org/10.1016/j.eswa.2011.12.023.

[16] L.H.A. Monteiro, D.M. Gandini, P.H.T. Schimit, The influence of immune individuals in disease spread evaluated by cellular automaton and genetic algorithm, Comput. Methods Progr. Biomed. 196 (2020), 105707, https://doi.org/10.1016/j.cmpb.2020.105707.

[17] N. Ru, Y. Jian-hua, D. Shuai-qi, L. Yang-guang, Wave impedance inversion in coalfield based on immune genetic algorithm, Procedia Earth Planet. Sci. 1 (2009) 929–935, https://doi.org/10.1016/j.proeps.2009.09.144.

[18] A. Utamima, A comparative study of hybrid estimation distribution algorithms in solving the facility layout problem. Egypt, Inform J. 22 (2021) 505–513, https://doi.org/10.1016/j.eij.2021.04.002.

[19] Z.-Q. Zhang, B. Qian, R. Hu, H.-P. Jin, L. Wang, A matrix-cube-based estimation of distribution algorithm for the distributed assembly permutation flow-shop scheduling problem, Swarm Evol. Comput. 60 (2021), 100785, https://doi.org/10.1016/j.swevo.2020.100785.

[20] R. Pérez-Rodríguez, A. Hernández-Aguirre, A hybrid estimation of distribution algorithm for the vehicle routing problem with time windows, Comput. Ind. Eng. 130 (2019) 75–96, https://doi.org/10.1016/j.cie.2019.02.017.

[21] Z.-Q. Zhang, B. Qian, R. Hu, H.-P. Jin, L. Wang, J.-B. Yang, A matrix-cube-based estimation of distribution algorithm for blocking flow-shop scheduling problem with sequence-dependent setup times, Expert Syst. Appl. 205 (2022), 117602, https://doi.org/10.1016/j.eswa.2022.117602.

[22] O.G. Toledano-López, J. Madera, H. González, A. Simón-Cuevas, A hybrid method based on estimation of distribution algorithms to train convolutional neural networks for text categorization, Pattern Recogn. Lett. 160 (2022) 105–111, https://doi.org/10.1016/j.patrec.2022.06.008.

[23] Z.-Q. Zhang, R. Hu, B. Qian, H.-P. Jin, L. Wang, J.-B. Yang, A matrix cube-based estimation of distribution algorithm for the energy-efficient distributed assembly permutation flow-shop scheduling problem, Expert Syst. Appl. 194 (2022), 116484, https://doi.org/10.1016/j.eswa.2021.116484.

[24] C. Wu, L. Wang, J. Wang, A path relinking enhanced estimation of distribution algorithm for direct acyclic graph task scheduling problem, Knowl.-Based Syst. 228 (2021), 107255, https://doi.org/10.1016/j.knosys.2021.107255.

[25] Q. Dang, W. Gao, M. Gong, An efficient mixture sampling model for Gaussian estimation of distribution algorithm, Inf. Sci. 608 (2022) 1157–1182, https://doi.org/10.1016/j.ins.2022.07.016.

[26] J. Zhao, Z. Yu, X. Yang, Z. Gao, W. Liu, Short term traffic flow prediction of expressway service area based on STL-OMS, Phys. Stat. Mech. Its Appl. 595 (2022), 126937, https://doi.org/10.1016/j.physa.2022.126937.

[27] H. Apaydin, M. Taghi Sattari, K. Falsafian, R. Prasad, Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions, J. Hydrol. 600 (2021), 126506, https://doi.org/10.1016/j.jhydrol.2021.126506.

[28] D. Wu, Y. Wang, L. Li, P. Lu, S. Liu, C. Dai, Y. Pan, Z. Zhang, Z. Lin, L. Yang, Demand response ability evaluation based on seasonal and trend decomposition using LOESS and S–G filtering algorithms, Energy Rep., ICPE 2021 - 2nd Int. Conf. Power Eng. 8 (2022) 292–299, https://doi.org/10.1016/j.egyr.2022.02.139.