

Research on gas concentration identification based on sparrow search algorithm optimization SVR

Yuanman Zhang¹, Yanan Zou^{2,*} and Qingyun Wu¹

¹School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin City 132022, Jilin Province, China

²School of Science, Jilin Institute of Chemical Technology, Jilin City 132022, Jilin Province, China

*E-mail: zouyanan@jiclt.edu.cn

Received for publication
October 13, 2024.

Abstract

To address the challenge of quantitatively identifying mixed gases, we developed a gas concentration identification algorithm based on the sparrow search algorithm (SSA) and optimized support vector regression (SVR). The Tent chaotic mapping operator is employed to initialize the population, enhancing population diversity, and improving the algorithm's global search capability. By optimizing SVR parameters with SSA, we propose an enhanced TSSA-SVR model. Evaluated on mixed gas datasets, TSSA-SVR achieves a prediction accuracy of 94.47%, outperforming comparative algorithms such as Genetic Algorithm (GA)-SVR and PSO-SVR, while demonstrating improved convergence compared to the baseline SSA-SVR. The experimental results demonstrate significant performance enhancements, offering an effective solution for precise gas concentration identification in complex environments.

Keywords

gas concentration prediction, sparrow search algorithm, support vector regression, electronic nose, gas sensor array

1. Introduction

In recent years, with the continuous expansion of industrial production worldwide, the detection of gas composition and pollution prevention have become growing concerns for both governments and individuals globally. Industrial production processes inevitably release a large amount of gas. Once the concentration exceeds a certain threshold, production lines may be forced to shut down, not only polluting the environment and disrupting work but also posing a serious threat to human life and safety. Some furniture after interior decoration will continue to release formaldehyde gas; long-term inhalation of excessive formaldehyde gas will endanger human life and health [1]. The chemical mixing of some gaseous constituents in the air with exhaust gases from vehicles can lead to the formation of acid rain, which accelerates the corrosion of vehicles and road infrastructure,

increases the risk of traffic accidents, and poses a threat to human health. It can be seen that harmful gases have a negative impact on work production and daily life. For such consideration, it is of great significance to identify and analyze gases qualitatively and quantitatively [2].

The research on gas recognition is constantly advancing, and the identification methods are constantly being iterated and upgraded. Humans perceive and analyze information through five senses: sight, hearing, touch, taste, and smell, enabling them to make corresponding reactions and decisions. With the continuous development of science and technology, the electronic bionic technology produced by this kind of biological sensing technology is increasingly appearing in the public view. Initially, humans relied on smell to analyze and process gas information. However, this method is subjective and heavily dependent on the individual tester's sense of

smell, which limits its reliability and consistency. Later, the use of expensive chemical analysis instruments such as mass spectrometry or chromatography for identification methods [3], this method requires not only the number and quality of operators but also cumbersome operation, time-consuming, and not cost-effective.

Nowadays, with the development of semiconductor technology, sensors integrated with advanced Micro-Electro-Mechanical Systems (MEMS) technology have been used to address the issues of high-power consumption and large size in earlier electronic nose systems [4]. These sensors offer advantages such as fast response, high precision, and long service life. This type of electronic nose system, which combines a gas sensor array with pattern recognition algorithms, is more agile and cost-effective, and it has promising applications in the detection of hazardous volatile organic compounds.

For example, harmful gases such as formaldehyde, once detected by humans, are transmitted to olfactory receptors and then to the brain through olfactory cells for recognition. Similarly, a bionic electronic device “smells” the gas through a sensor array and processes the signals in a system analogous to the human brain for classification, recognition, and prediction. Since 1964, researchers have discovered that the anatomy of the olfactory region is limited, and the principles underlying gas concentration changes were not well understood. Through experiments, it was found that olfactory simulation using electrodes could replicate these processes [5]. Then, in 1982, the original electrode was improved and an electronic nose device equipped with a semiconductor sensor was designed [6], which proved to be a device with a certain odor identification ability. By 1984, Bartlett, aiming at the cross-sensitivity phenomenon of semiconductor sensors, composed an array of various gas sensors to identify gas [7], which fully proved that this is a small portable device capable of providing detection capabilities and provided a theoretical basis for electronic nose technology. In 1994, Gardner gave the electronic nose a professional definition: an instrument consisting of a sensor array and a pattern recognition system, and since then, electronic nose systems have embarked on the path of specialization [8].

The electronic nose system comprises three main modules: a sensor array, signal acquisition and conversion, and a pattern recognition algorithm [9]. The gas mixture to be detected is first collected by a gas sensor array. The advantage of using a sensor array for data collection is that different sensors exhibit distinct responses to various gases, facilitating better

handling of mixed gas environments. The type of gas sensor employed in the research is a metal oxide semiconductor (MOS) based gas sensor. The MOS gas sensor detects gas concentrations through chemical reactions on the surface of its sensitive material, such as tin dioxide (SnO_2), which interacts with target gases and causes changes in electrical resistance. These changes can be digitized, forming the original dataset. The pre-processing module is then used to normalize the collected original dataset, remove noise interference, and clean the data effectively. Finally, a suitable algorithm is used in the pattern recognition module to identify and detect the target gas.

The electronic nose system is a device that employs a combination of gas sensors and recognition algorithms to sense and classify odors. Traditionally, research in this field has focused primarily on improving sensor materials and pattern recognition methods. For example, many studies have enhanced the performance of gas sensors through composite materials, advanced preparation techniques, and doping. Doping Pt or Nb into zinc oxide, for instance, has been shown to significantly improve sensor responses [10]. However, such hardware improvements face limitations due to factors such as operating temperature, material cost, and synthesis complexity. These challenges underscore the importance of integrating advanced pattern recognition algorithms to enhance the accuracy of electronic nose systems in gas recognition. Recent advances in algorithmic approaches have brought significant improvements. Li et al. [11] developed a one-dimensional convolutional neural network (CNN) capable of adaptively extracting features from response signals, achieving a remarkable accuracy of 99.98% in mixed gas identification. Similarly, Wang [12] integrated particle swarm optimization (PSO) and artificial bee colony (ABC) algorithms with extreme learning machines (ELM) to detect six pollutants, achieving a high correlation coefficient of 0.99 in monitoring results. These successes demonstrate the potential of swarm intelligence algorithms in addressing complex gas recognition problems. Among swarm intelligence methods, the sparrow search algorithm (SSA) has emerged as a particularly promising approach due to its bio-inspired optimization mechanism. Hu et al. [13] applied SSA to optimize the weights and thresholds of a Back Propagation (BP) neural network, solving the slow convergence issues in traditional methods and improving air quality index prediction. Building on such advancements, recent studies have further refined SSA for specific applications. Zhu et al. [14] introduced an improved SSA

(ISSA) combined with least-squares support vector machines (LSSVM) for multi-component gas concentration prediction. The ISSA significantly enhanced prediction accuracy compared to PSO and traditional SSA methods. Kupin and Kosei [15] provided an in-depth review of swarm intelligence algorithms, emphasizing their application in solving complex optimization problems. Future research directions, including hybrid approaches combining swarm intelligence with other artificial intelligence methods, were also discussed. Wang and Bi [16] proposed a support vector regression (SVR) model optimized by PSO to address challenges in greenhouse gas detection accuracy and parameter selection. This work demonstrates the potential of PSO in fine-tuning model parameters for enhanced predictive performance.

Recent advances in gas concentration prediction have emphasized the integration of machine learning algorithms with optimization techniques. Therefore, in this paper, SSA is used to optimize the training of SVR. At the same time, in order to avoid the algorithm easily falling into local optimality, Tent mapping is added to optimize the TSSA-SVR model with high prediction accuracy and strong global search ability. Finally, the model is applied to the sample data of gas concentration identification, which provides a feasible method for gas concentration identification.

II. Classic sparrow algorithm and improvement strategy

a. Classic sparrow algorithm

The sparrow is a living creature mainly living in groups, with quick-witted, bold and careful, and high alertness characteristics. In 2020, Xue and Shen [17] proposed a SSA inspired by the foraging behavior of sparrow groups.

The division of labor in the foraging activities of the sparrow group is clear, and according to the nature of work, it can be divided into three categories: seeker, follower, and alert. The hunter is responsible for searching for food, which is a survival skill that every sparrow possesses, but the finder is present in a certain percentage of foraging activities. The seeker has a high fitness in iterative search and a wide range of search capabilities, and its position update is shown in Eq. (1). $iter_{max}$ indicates the maximum number of iterations, α is a uniform random number in (0, 1), Q is a random number that follows a standard normal distribution, L represents an all-1 matrix of $1 \times d$, where d is the dimension of the objective function, and $R_2 \in [0, 1]$ and $ST \in [0.5, 1]$ indicate alarm value and alarm

threshold, respectively. Only when the alarm value is less than the alarm threshold, the finder can find food more at ease and increase the probability of finding food.

$$x_{u,v}^{t+1} = \begin{cases} x_{u,v}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right), & R_2 < ST \\ x_{u,v}^t + QL, & R_2 \geq ST \end{cases} \quad (1)$$

The follower will follow the seeker with the best fitness to forage in order to obtain better food, and its position is updated as shown in Eq. (2). x_p^{t+1} shows the best position of the discoverer on the $(t + 1)$ iteration, x_{worst}^t shows the worst position in the population at the t iteration, and A is a $1 \times d$ matrix with only 1 or -1, $A^+ = A^T (AA^T)^{-1}$. When $i > n/2$, the follower [3] has poor fitness and needs to choose other places to forage.

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{x_{worst}^t - x_{i,j}^t}{\alpha \cdot iter_{max}}\right), & i > n/2 \\ x_p^{t+1} + |x_{i,j}^t - x_p^{t+1}| \cdot A^+ \cdot L, & i \leq n/2 \end{cases} \quad (2)$$

Alerters need to be aware of the movement of predators at all times [18]. Also in proportion to the presence, once there is a danger to quickly remind the whole population to fly to a safe area, its position update is shown in Eq. (3). Where, x_{best}^t represents the optimal position in the population in the t iteration, β is a normally distributed random number with a mean of 0 and a variance of 1, and ε is a constant that approaches 0, $k \in [-1, 1]$,

$$x_{i,j}^{t+1} = \begin{cases} x_{best}^t + \beta \cdot |x_{i,j}^t - x_{best}^t|, & f_i \neq f_g \\ x_{i,j}^t + k \frac{|x_{i,j}^t - x_{worst}^t|}{(f_i - f_w) + \varepsilon}, & f_i = f_g \end{cases} \quad (3)$$

f_g is the optimal fitness value of sparrow population at present, and f_w is the worst fitness value of sparrow population. When $f_i \neq f_g$, sparrows are in the optimal foraging position of the population and can forage around confidently. When $f_i = f_g$, the sparrow is in the optimal position to easily detect the danger signal, and will give up foraging and go to the safe position once detected [19].

During the iteration process, as long as the sparrow's new position is better than the previous position, the position is updated. The fitness value is updated until the best position and the best fitness are found. During this period, the sparrow's identity

was also constantly updated and alternating. If well adapted, each sparrow can become a finder, but the finder-to-entrant ratio is constant.

b. Tent map improved sparrow algorithm

As a non-linear phenomenon, chaos has some random and irregular variables, which increase the uncertainty and non-repeatability. However, it is also true that chaos variables have the characteristics of randomness and ergodicity, which directly affect the population diversity of the algorithm and improve the global search ability of the algorithm. When applied to optimization problems, it can greatly improve the optimization ability of the algorithm [20]. Using chaotic mapping to initialize the population can not only keep the diversity of the population but also jump out of the local optimal and improve the global search ability.

The Logistic mapping function operator and the Tent chaotic mapping function operator are the most used chaotic operators at present. The Logistic mapping function operator is a valley-type distribution state with two ends high and middle low, which has the phenomenon of traversal non-uniformity, and its optimization speed is lower than that of Tent mapping. Moreover, the Tent mapping function operator presents a more uniform distribution state and can generate a better uniform sequence. In contrast, mapping works better using Tent chaotic [21]. The expression for Tent mapping is shown in Eq. (4). z_{i+1} indicates the result of the $i + 1$ iteration, z_i indicates the number of iterations, and λ represents the control parameter when λ is in the range [2,4], the system is in a chaotic state, and the value of λ in this paper is 2.

$$z_{i+1} = \begin{cases} \lambda z_i, & 0 \leq z_i \leq \frac{1}{2} \\ \lambda(1 - z_i), & \frac{1}{2} < z_i \leq 1 \end{cases} \quad (4)$$

The quality of the initial sparrows in the SSA has a certain influence on the result. After a certain number of iterations, the sparrows in the group gradually approach an optimal solution. By continuously iterating the Tent function, we obtain a sequence of pseudo-random numbers $\{x_1, x_2 \dots x_T\}$, which will be used to initialize the initial positions of the sparrows. The output of the Tent function is a random number within the interval [0, 1]. We need to map it to the actual search space of the optimization problem. Suppose the search space of the optimization problem is $[LB_j, UB_j]$ where LB_j and UB_j represent the lower and upper bounds of the j -th dimension, respectively.

Then, the j -th dimension position of the sparrow can be calculated by the following equation:

$$P_{i,j} = LB_j + X_T \cdot (UB_j - LB_j). \quad (5)$$

Using Eq. (5), the variable values generated by the Tent function are mapped to individual sparrows, achieving the initialization of the sparrow population.

By introducing the Tent chaotic function to initialize the population of SSA, this approach generates better uniform sequences, helps capture higher-quality populations, lays the foundation for subsequent research, improves the algorithm's global search capability [22–23], and enhances the algorithm's performance.

c. Principle of support vector machine (SVM) algorithm

Before neural networks were widely used, the SVM algorithm was widely used in the field of pattern recognition [24], and it is a supervised learning algorithm based on statistical learning theory. It converts linearly indivisible problems in a low-dimensional space to a high-dimensional space to find an optimal solution that can separate different types of data, so that the interval between positive and negative samples is maximum. SVR is an important application branch of SVM, which has a complete theoretical foundation and good non-linear prediction ability, and is a supervised machine learning method. The basic idea is no longer to find an optimal classification surface to separate samples but to find an optimal classification surface to minimize the error between all training samples and the optimal classification surface. The regression task is realized by finding an optimal hyperplane that minimizes the distance between sample points above and below the hyperplane. For the problem of linear indivisibility, the SVR algorithm maps the sample data to the high-dimensional feature space by kernel function [25], converts the non-linear operation into linear operation, then finds the linear decision function, and makes regression prediction through the decision function.

In the SVR algorithm, the selection of parameter values has a direct influence on the prediction accuracy and generalization ability of the model. Traditional SVR parameter optimization methods include cross-validation, gradient descent, and grid search [26]. Each of the three has its own advantages and disadvantages: cross-validation technology is computationally heavy, and it is difficult to select more than two parameters. The gradient descent method has faster computation time and convergence speed,

but it requires the objective function to be differentiable to the parameters. When there are many parameters or a wide range of parameter values, the grid search method requires a large amount of computation and takes a long time [27]. Therefore, to improve the learning ability and generalization ability of the SVR model, we can start by optimizing the key parameters and superimposing various optimization algorithms to achieve great progress in the application of the algorithm.

III. Tent-SSA-SVR algorithm

a. Algorithmic thought

The improvement of the SVR algorithm focuses on two major parameters that have a great impact on the performance of the SVR algorithm [28]. The first is the penalty factor C , which represents the relationship between classification complexity and classification accuracy. The smaller C is, the lighter the penalty for error will be, which will easily lead to model under-fitting. On the contrary, the larger C is, the smaller the tolerance of the model to errors, which is easy to make the model overlearn the training data, resulting in overfitting. The second key parameter that has a great influence on the performance is the kernel function g , which controls the radial effect range of the

model. The SSA has the advantages of fast convergence speed and little influence by parameter setting, so it has certain advantages in finding the optimal value of these two parameters. At the same time, in order to reduce the probability of SSA falling into local optimal in the process of parameter optimization, the Tent chaotic function operator is used to initialize the population, forming a more uniform distribution of the initial population, improving the global optimal performance and increasing the accuracy of gas concentration prediction.

b. Tent-SSA-SVR algorithm design for gas concentration prediction

According to this algorithm idea, the optimal parameters C and g are found to form the TSSA-SVR algorithm model to predict the concentration of mixed gas. The flow chart is shown in Figure 1, and the specific steps are described as follows.

Step 1: Normalize the original data collected by the gas sensor to prevent slow convergence caused by uneven data. The normalized data are divided into a 70% training set and a 30% test set.

Step 2: Use Tent Chaos to initialize the population position and select the individuals with the best fitness as the initial population.

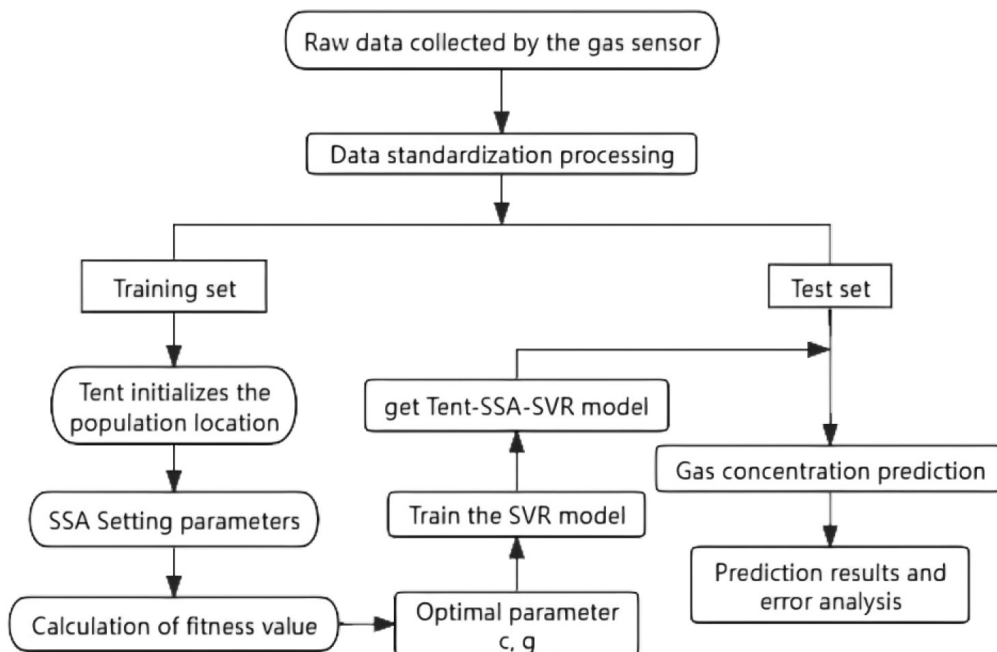


Figure 1: Flow chart of the gas concentration prediction algorithm. SSA, sparrow search algorithm; SVR, support vector regression.

Step 3: Initialize related parameters: population size, maximum number of iterations, number of discoverers, number of early warning sparrows, etc.

Step 4: Calculate the fitness value of each sparrow, select the current optimal position and optimal fitness value, and select the current worst position and worst fitness value. The sparrows with the highest fitness were selected as the seekers and the others as the followers. The location of the finder was constantly updated according to the warning and safety values.

Step 5: Compare the optimal position and fitness according to the current state of the population. When the maximum number of iterations is reached, the optimal solution is output; otherwise, the iteration continues. The optimal parameter values C and g are obtained by iterative optimization.

Step 6: The optimal parameters C and g are substituted into the SVR model for training, and the TSSA-SVR algorithm model is obtained.

Step 7: Apply the TSSA-SVR algorithm model to the mixed gas data to analyze and predict the concentration value of the mixed gas.

IV. Simulation experiment and result analysis

a. Experimental environment and dataset

In order to verify the feasibility of this optimization model, the dataset "Gas sensor array in Dynamic Gas mixture" was selected from the University of California, Irvine (UCI) machine learning library. In the process of data collection, the sample data were made as close as possible to the real environment, that is, the sample data were made to have the characteristics of irregular distribution in the real environment. Use 70% of the dataset as the training set and 30% as the test set. The dataset contains 16 chemical sensors of four different types: Tin Oxide Gas Sensor (TGS)-2602, TGS-2602, TGS-2600, TGS-2600, TGS-2610, TGS-2620, TGS-2620, TGS-2602, TGS-2602, TGS-2600, TGS-2600, TGS-2610, TGS-2610, TGS-2620, TGS-2620, and TGS-2610. The gas sample is mainly composed of methane and ethylene in the air, with ethylene concentrations ranging from 0 ppm to 20 ppm and methane concentrations ranging from 0 ppm to 300 ppm. The concentration of the gas sample changes randomly during the experimental measurement, and 16 sensor arrays collect

data continuously. More details on the dataset can be found in Fonollosa et al. [29]. *Experimental environment:* The software used was Python and the operating system used was Windows 10.

In the experiment, the original data need to be pre-processed, and the pre-processed data are normalized to 0–1, which reduce the errors caused by the different dimensions of the data and improve the computational efficiency of the regression model fitting process [30]. The normalized calculation method is shown in Eq. (6).

$$Y = \frac{x - \text{Min}}{\text{Max} - \text{Min}} \quad (6)$$

Next, principal component analysis (PCA) is used to reduce the dimensionality and extract features from the sensor array data. This method not only reduces the data dimensionality but also maximizes the retention of key information in high-dimensional data and solves the problems of dimensional inconsistency and redundant information. The covariance matrix of the normalized data is calculated as:

$$C = \frac{1}{n-1} X^T X. \quad (7)$$

The principal components were extracted by performing eigenvalue decomposition of the covariance matrix. The contribution rate of each eigenvalue p_i was calculated as:

$$p_i = \frac{\lambda_i}{\sum \lambda_i}. \quad (8)$$

Based on the cumulative contribution rate, the top four principal components were selected, accounting for > 99% of the total variance. These principal components serve as the reduced feature set, significantly reducing the dimensionality of the data while preserving the essential information. After dimensionality reduction, the dataset was randomly divided into training and test sets in a 7:3 ratio. This approach retains the primary features of the sensor data while reducing computational complexity and minimizing the impact of redundant information.

b. Parameter setting and analysis of experimental results

Table 1 shows the initial parameter settings when SSA is used to optimize the two key parameters of SVM.

To verify the effectiveness of TSSA optimization of SVR parameters, the proposed algorithm was compared with traditional SVR prediction and SSA optimization SVR algorithms. The parameter settings of each model are shown in Table 2.

c. Analysis of experimental results

After the parameters are set, the experimental verification is carried out according to the algorithm flow chart, and the five-fold cross-verification is adopted. The mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2) were used to evaluate the prediction accuracy of each model. The greater the average absolute error value, the greater the error; the RMSE judgment is the same as the average absolute error, the greater the value, the greater the error; the coefficient of determination R^2 is used to reflect the accuracy of fitting data to the regression model, and its value ranges from [0, 1]. The closer the value is to 1, the better the model is. The calculation formula for the three evaluation indicators [31] is as follows:

$$MAE = \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i| \quad (9)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (11)$$

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i \quad (12)$$

$\bar{y}$ is the average of the predicted values of my test samples. The error results of the three algorithms for quantitative identification of mixed gases are shown in Table 3.

As shown in Table 3, the MAE and RMSE of the SVR algorithm optimized by SSA for quantitative recognition of gas mixtures are 1.485 and 1.982. The MAE is 0.425 and the RMSE is 0.286 after the SSA optimized by the Tent chaotic function operator is combined with SVM. First, compared with the single algorithm model that only uses SVR, the evaluation criterion of the determination coefficient is improved by 0.801, and the evaluation criterion of the determination coefficient is improved by 0.521 compared with the SSA-SVR model, which is sufficient to prove that compared with the traditional machine learning algorithm, the proposed algorithm has greatly improved the test

Table 1: Initial parameter table

Parameter	Value
Maximum iterations	200
Population size	100
Optimization parameter number	2
Upper limit of optimization parameter	2^{-10}
Lower limit of optimization parameter	2^{10}
Cross-check the number of folds	5

Table 2: Model parameter settings

Algorithm	Parameter	Value range	Optimum value
SVR	C	/	10
	g		0.5
SSA-SVR	Alarm value	[0.5, 1]	0.8
	Safety threshold	[0, 1]	0.2
TSSA-SVR	Alarm value	[0.5, 1]	0.8
	Safety threshold	[0, 1]	0.2

SSA, sparrow search algorithm; SVR, support vector regression.

Table 3: Quantitative identification results of mixed gases

Model	RMSE	MAE	R^2	Acc (%)
SVR	2.626	2.336	0.095	81.80
SSA-SVR	1.982	1.485	0.375	87.43
GA-SVR	1.892	1.367	0.482	89.67
PSO-SVR	1.765	1.220	0.512	90.36
TSSA-SVR	0.286	0.425	0.896	94.47

MAE, mean absolute error; PSO, particle swarm optimization; RMSE, root means square error; SSA, sparrow search algorithm; SVR, support vector regression.

results and has practical application value in the field of gas identification. Second, the closer the coefficient of determination (R^2) is to 1, the better the pattern

recognition performance. As can be seen, the value closest to 1 in Table 3 is 0.896, which corresponds to the TSSA-SVR algorithm model. This algorithm has some advantages compared with other algorithms.

To better evaluate the optimization capability of the proposed algorithm in improving model prediction accuracy, we employed an intuitive curve-based analysis. Figure 2 shows the fitness change curves of the two types. It can be seen that the algorithms converge gradually with the increase in the number

of iterations, and the fitness curve of the Tent-SSA algorithm is gentler, indicating that the Tent-SSA algorithm has a higher impact on the convergence performance.

In order to further intuitively feel the improvement of recognition accuracy, as shown in Figure 3, the two algorithm models optimized by chaotic sequence and those not optimized by chaotic sequence were compared together to evaluate the absolute error value of the gas samples not obtained by the standard.

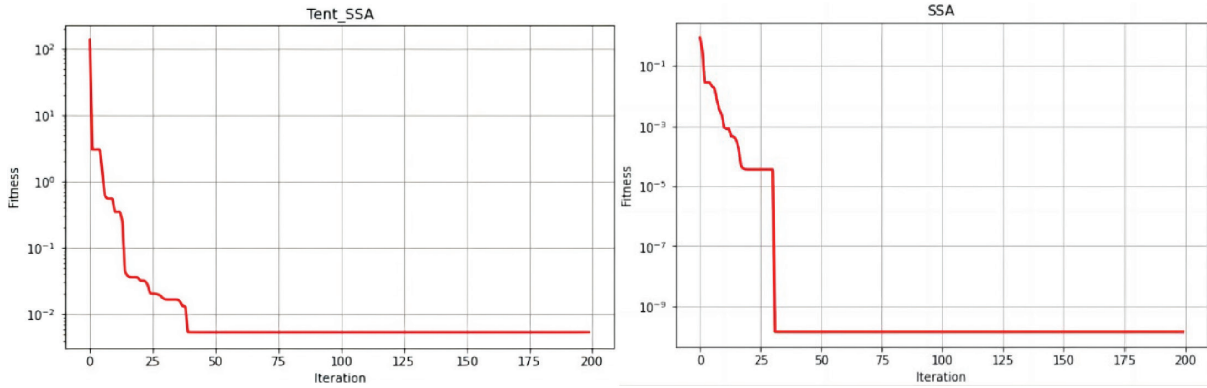


Figure 2: Fitness change curves of two different algorithms. SSA, sparrow search algorithm.

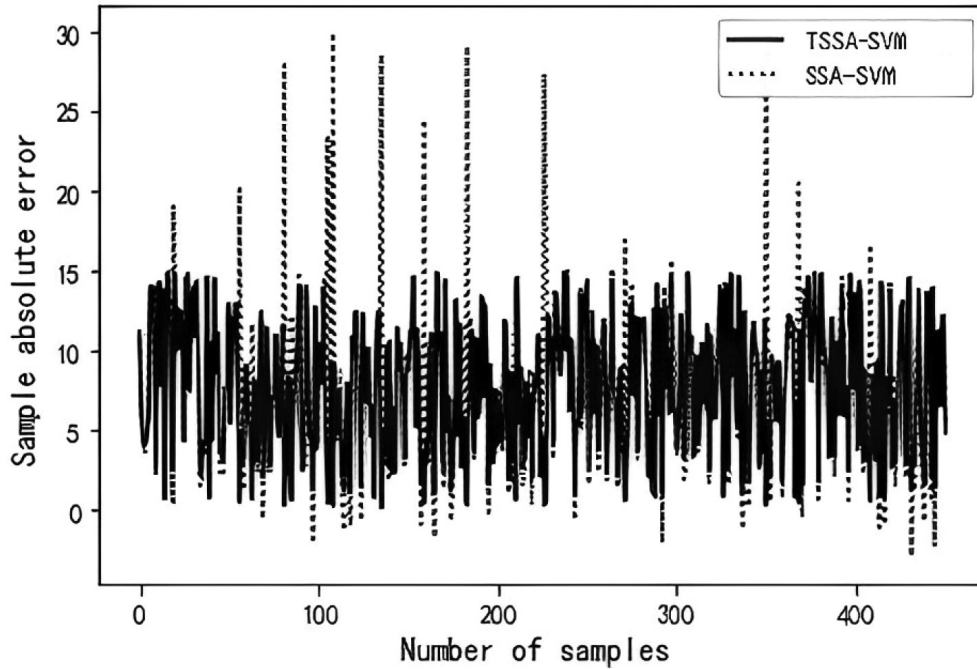


Figure 3: The absolute error of the two algorithm models. SSA, sparrow search algorithm; SVM, support vector machine.

As shown in Figure 3, the absolute error of the TSSA-SVR algorithm is significantly lower than that of the SSA-SVR on most samples, and the overall error curve is smoother and less volatile. This indicates that the algorithm has higher accuracy and robustness in the prediction process, which avoids large ups and downs in the prediction results and thus improves the stability of the overall recognition effect. It can be seen that the algorithm model proposed in this paper can improve the accuracy of the concentration identification of ethylene and methane mixed gas in the air, and the degree of fitting of the algorithm is moderate, which has a certain application value.

V. Conclusions

At present, many scholars have done a lot of research on the pattern recognition algorithm of gas recognition, and many excellent recognition models have appeared. Inspired by the general application advantage of the SVM algorithm, this paper uses an optimization algorithm to design an improved pattern recognition algorithm. This study proposes a gas concentration identification method based on a Tent chaotic mapping-enhanced sparrow search algorithm (TSSA) optimized SVR. By integrating Tent chaotic sequences to initialize the sparrow population, the global search capability of the algorithm is enhanced, which in turn improves the parameter optimization efficiency of SVR. Experimental results demonstrate the superiority of the TSSA-SVR model in both gas concentration predictions. Key findings are summarized below:

- (1) TSSA-SVR outperforms traditional SVR, SSA-SVR, GA-SVR, and PSO-SVR across all evaluation metrics. It achieves an RMSE of 0.286, MAE of 0.425, MAE of 0.425, R^2 of 0.896, and classification accuracy of 94.47%, significantly surpassing other models.
- (2) The Tent chaotic mapping enhances the global search capability of the SSA, enabling the TSSA-SVR to avoid local optima and achieve faster convergence. Through fitness curve analysis, TSSA-SVR demonstrates smoother and more stable convergence compared with the traditional SSA-SVR.
- (3) In terms of data acquisition in this paper, with the long-term use of sensors, sensor drift will inevitably occur, resulting in data distortion and other phenomena, affecting data analysis. Next, we will focus on this problem and reduce the impact of

drift on subsequent gas identification from two perspectives, that is, software design and hardware improvement.

Funding

This work was financially supported by the National Science Foundation of China (grant no. 62201225).

References

- [1] C. Jung, N. Mahmoud, N. Qassimi, and G. Elsamanoudy, "Preliminary Study on the Emission Dynamics of TVOC and Formaldehyde in Homes with Eco-Friendly Materials: Beyond Green Building," *Buildings*, Vol. 13, No. 11, 2023, pp. 2847.
- [2] M. Li, J. He, R. Zhou, et al., "Research on Prediction Model of Mixed Gas Concentration Based on CNN-LSTM Network," *Proceedings of the 3rd International Conference on Advanced Information Science and System*, 2021, pp. 1-5.
- [3] C. Zhao, J. Ma, W. Jia, et al., "An Apple Fungal Infection Detection Model Based on BPNN Optimized by Sparrow Search Algorithm," *Biosensors*, Vol. 12, No. 9, 2022, pp. 692.
- [4] S. Mu, W.F. Shen, and D.W. Lu, "Research Progress of Electronic Nose Technology and Its Application," *Materials Review*, 2024, pp. 1-34.
- [5] W.F. Wilkens and J.D. Hartman, "An Electronic Analog for the Olfactory Processes," *Journal of Food Science*, Vol. 29, No. 3, 1964, pp. 372-378.
- [6] K. Persaud and G. Dodd, "Analysis of Discrimination Mechanisms in the Mammalian Olfactory System Using a Model Nose," *Nature*, Vol. 299, No. 5881, 1982, pp. 352-355.
- [7] S. Zaromb and J.R. Stetter, "Theoretical Basis for Identification and Measurement of Air Contaminants Using an Array of Sensors Having Partly Overlapping Selectivities," *Sensors and Actuators*, Vol. 6, No. 4, 1984, pp. 225-243.
- [8] J.W. Gardner and P.N. Bartlett, "A Brief History of Electronic Noses," *Sensors and Actuators B: Chemical*, Vol. 18, Nos. 1-3, 1994, pp. 210-211.
- [9] C. Zhan, J. He, and M. Pan, "Component Analysis of Gas Mixture Based on One-Dimensional Convolutional Neural Network," *Sensors*, Vol. 21, No. 2, 2021, pp. 347.
- [10] Z. Li, Z. Yao, A. Haidry, T. Plecenik, B. Grančič, T. Roch, M. Gregor, and A. Plecenik, "The Effect of Nb Doping on Hydrogen Gas Sensing Properties of Capacitor-Like Pt/

Nb-TiO₂/Pt Hydrogen Gas Sensors,” *Journal of Alloys and Compounds*, Vol. 803, 2019, pp. 225-233.

[11] P. Li, Y. Xu, J. Yang, et al., “Research on Gas Recognition Method Based on One-Dimensional Convolutional Neural Network,” *Electronic Devices*, Vol. 45, No. 3, 2022, pp. 645-650.

[12] J. Wang, Y. Tao, and Z. Liang, “Electronic Nose Gas Concentration Detection Based on Improved Extreme Learning Machine,” *Application of Electronic Technology*, Vol. 47, No. 10, 2019, pp. 63-67.

[13] Q. Hu, S. Gong, and Z. Hu, “Air Quality Index Prediction Based on Improved Sparrow Search Algorithm,” *Guangxi Science*, Vol. 29, No. 4, 2022, pp. 642-651.

[14] Z. Zhu, B. Tian, X. Fan, M. Zeng, and Z. Yang, “Concentration Prediction of Multi-component Gases Based on Improved Sparrow Search Algorithm,” *Journal of Physics: Conference Series*, Vol. 2650, 2023.

[15] Kupin and M. P. Kosei, “Analysis of Swarm Intelligence Algorithms,” *System Technologies*, 2024.

[16] X. Wang and M. Bi, “Greenhouse Gas Prediction Method Based on Particle Swarm Optimized SVR,” 12511, Vol. 12511, 2023, pp. 1251126-1251126-6.

[17] J. Xue and B. Shen, “A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm,” *Systems Science & Control Engineering*, Vol. 8, No. 1, 2020, pp. 22-34.

[18] X. Lv, X. Mu, and J. Zhang, “Chaotic Sparrow Search Optimization Algorithm,” *Journal of Beijing University of Aeronautics and Astronautics*, Vol. 47, No. 8, 2021, pp. 1712-1720.

[19] K. Meng, C. Chen, and B. Xin, “MSSSA: A Multi-Strategy Enhanced Sparrow Search Algorithm for Global Optimization,” *Frontiers of Information Technology & Electronic Engineering*, Vol. 23, No. 12, 2022, pp. 1828-1847.

[20] C. Xiang, H. Shi, et al., “Research on Bridge Damage Identification Method Based on Modal Frequency Strain Energy Entropy and Tent-SSA-BP Neural Network,” *Highway*, Vol. 68, No. 3, 2019, pp. 143-150.

[21] Z.Q. Bao, C. Lu, and S. Zhang, et al., “Research on Coke Quality Prediction Model Based on TSSA-SVR Model,” *China Mining*, Vol. 31, No. 6, 2022, pp. 86-92.

[22] J. Chen, Y. Fan, and X. Dai, “Research on Intelligent Vehicle Path Planning with Improved Sparrow Search Algorithm,” *Journal of Chongqing University of Technology (Natural Science)*, Vol. 37, No. 4, 2023, pp. 50-56.

[23] X. Wang and Q. Zhang, “Loss Prediction by Support Vector Machine with Improved Sparrow Search Algorithm,” *Science Technology and Engineering*, Vol. 22, No. 34, 2022, pp. 15115-15122.

[24] H. Zhang and Y. Han, “A New Mixed-Gas-Detection Method Based on a Support Vector Machine Optimized by a Sparrow Search Algorithm,” *Sensors*, Vol. 22, No. 22, 2022, pp. 8977.

[25] L. Zhang, T. Wang, and H. Zhou, “Research Progress of SVR Parameter Optimization Based on Swarm Intelligence Algorithm,” *Computer Engineering and Applications*, Vol. 57, No. 16, 2019, pp. 50-64.

[26] Y. Tang, Q. Dai, and M.Y. Yang, “Improved Sparrow Search Algorithm to Optimize SVM Outlier Detection,” *Computer Engineering and Science*, Vol. 45, No. 2, 2023, pp. 346-354.

[27] W. Song, S. Liu, X. Wang, et al., “An Improved Sparrow Search Algorithm,” *Proceedings of the 2020 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, IEEE, 2020, pp. 537-543.

[28] C. Zhang and S. Ding, “A Stochastic Configuration Network Based on Chaotic Sparrow Search Algorithm,” *Knowledge-Based Systems*, Vol. 220, 2021, pp. 106924.

[29] Fonollosa, J.; Sheik, S.; Huerta, R.; Marco, S. Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring. *Sensors Actuators B Chem.* 2015, 215, pp.618–629.

[30] Y. Zou and J. Lv, “Using Recurrent Neural Network to Optimize Electronic Nose System with Dimensionality Reduction,” *Electronics*, Vol. 9, No. 12, 2020, pp. 2205.

[31] W. Wojnowski, T. Majchrzak, T. Dymerski, et al., “Electronic Noses: Powerful Tools in Meat Quality Assessment,” *Meat Science*, Vol. 131, 2017, pp. 119-131.