

MULTI SENSOR DATA FUSION BASED ON LOCAL AND GLOBAL FOR ROOM TEMPERATURE MONITORING

Jing Li^{1,*}, Fei Yu²

¹School of Artificial Intelligence, Xiamen City University, China

²School of Information Science and Engineering, South East University, China

Email: lijing@xmcu.edu.cn

Abstract - Currently, data collection from a single sensor is no longer sufficient to meet people's information needs. To accurately predict the actual temperature, humidity, and lighting conditions in greenhouse environments and have a positive effect on vegetable cultivation, this study proposes a radial basis function neural network optimized by cuckoo search algorithm, and combines the optimized Dempster Shafer theory for greenhouse environment prediction. The optimized radial basis function of the improved cuckoo search algorithm converged at 10 iterations, and the recall rate finally converged to around 0.9. The optimized radial basis function of the improved cuckoo search algorithm was at the minimum level among the three error values, with an average reduction of 0.14, 0.25, and 0.24 compared to the other two algorithms. The humidity was reduced by an average of 0.25, 0.49, and 0.39, and the lighting was reduced by an average of 3, 27, and 2. After introducing the improved Dempster Shafer theory in the second example, the uncertainty of the final result decreased from 32.3% to 23.9%, while the output probability increased from 11.3% to 68.5%. Therefore, the radial basis function optimized by the improved cuckoo search algorithm has better prediction accuracy for various indicators in the greenhouse, while the error is small, which can significantly reduce uncertainty. This study provides a theoretical basis for the layout of greenhouse environmental monitoring equipment in the vegetable production process.

Keywords: Room temperature detection, Data fusion, Sensor data, Local fusion center, Global Fusion Center

1. Introduction

Nowadays, the quantity of functions of intelligent devices is gradually increasing, and data fusion technology based on multiple sensors is gradually emerging [1-2]. The core idea of multi-sensor data fusion technology is to collect corresponding data from different types of sensors and fuse them to obtain data types with new meanings [3-4]. Multi sensor data fusion is important in autonomous driving, LiDAR, and greenhouse. It can achieve accurate prediction of the environment and navigation capabilities through the multidimensional information obtained [5-6]. In practical applications, there are also some problems with multi-sensor data fusion, such as the need to consume a large amount of computation time for data, as well as differences in sampling frequency and accuracy between sensors [7-8]. Currently, scholars have conducted relevant research on multi-sensor data fusion. Fei et

al. proposed a multi-sensor data fusion based on machine learning and unmanned aerial vehicles for wheat crop yield prediction to obtain better wheat breeding decisions. This method demonstrated higher accuracy while also having smaller errors [9]. Senel et al. proposed a multi-sensor module based on cameras, radar, and LiDAR for environmental perception in intelligent transportation applications. Multi sensor methods could detect potential faults with computation time much less than 10 milliseconds [10]. Liu et al. proposed a multi-sensor fusion algorithm to improve the accuracy of navigation for unmanned surface vehicles, to address their susceptibility to environmental influences. This method had a small difference between the error and the true value, and improved the accuracy and reliability of navigation [11]. Cao proposed a multi-sensor fusion diagnostic method for milling machine vibration to collect real-time cutting product information for system monitoring. This

method established a deep belief network, which improved its accuracy by 23% and achieved an accuracy value of up to 98.7% [12].

In the above study, although multi-sensor fusion achieved good results, the fusion effect of the system was not evaluated by combining local and global fusion centers. Therefore, the study proposes a multi-sensor data fusion based on local and global methods for room temperature monitoring. The innovation lies in the introduction of an optimized Cuckoo Search (CS) combined with radial basis functions to correct relevant outliers, while optimizing the D-S (Dempster Shafer) theory and applying it to the global fusion center.

This study aims to integrate and analyze environmental factors in greenhouses through multi-sensor fusion technology, and evaluate environmental information to assist vegetable growth.

2. Methods and Materials

2.1 Construction of Local Fusion Model Based on CS-RBF

The temperature, light, and humidity in the greenhouse environment play an important role in crop growth. When collecting temperature, humidity, and light inside the greenhouse through multiple sensors, it is first necessary to establish a local fusion center. This study divides it into abnormal data detection, error data adjustment based on radial basis function (RBF), and weighted average, but this method has a strong dependence on algorithm parameters. Therefore, the study introduces CS to update parameters. When using multiple sensors to collect data in a greenhouse, it is usually necessary to layout the sensors, as shown in Figure 1.

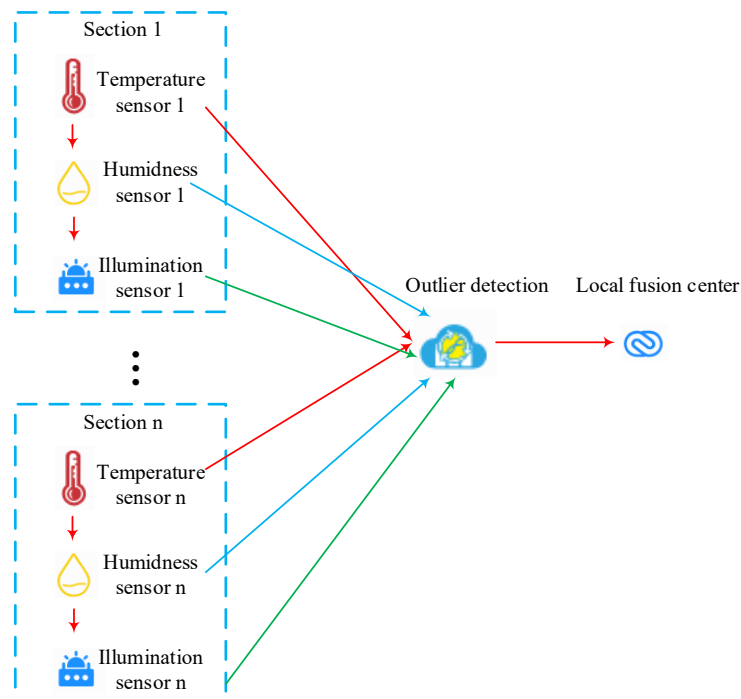


Figure 1: Schematic diagram of sensor area division

In Figure 1, after data is collected by sensors in multiple regions, it enters the classification fusion center through outlier detection. In the greenhouse environment, uncertain factors lead to sensors collecting inappropriate data, which in turn affects owners' decision-making. Common anomaly data detection methods include Dixon criterion, Grubbs criterion, and Rayda criterion [13]. Due to the fact that the above criteria require data samples to follow a normal distribution, which does not meet the relevant data types in greenhouse environments, the study introduces box plots to detect abnormal data. The relevant definitions of box plots are shown in Figure 2.

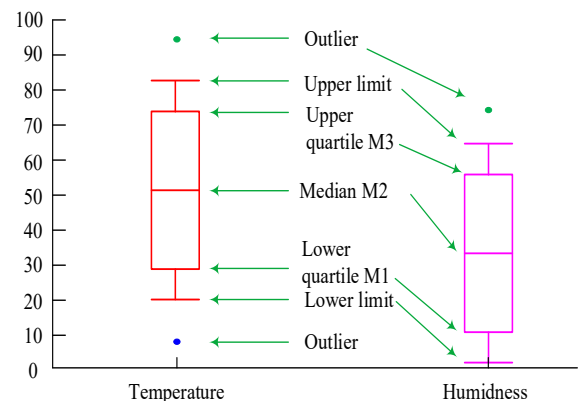


Figure 2: Box diagram structure diagram

In Figure 2, the box diagram mainly consists of five parts. The temperature type is defined to exist in the arranged data $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, so the relevant location information is shown in equation (1).

$$\begin{cases} n_1 = \frac{(n+1)}{4} = 1.75 \\ M_1 = x_1 + (x_2 - x_1)(n_1 - |n_1|) \\ n_2 = \frac{3(n+1)}{2} = 3.5 \\ M_2 = x_3 + (x_4 - x_3)(n_2 - |n_2|) \\ n_3 = \frac{3(n+1)}{4} = 5.25 \\ M_3 = x_5 + (x_6 - x_5)(n_3 - |n_3|) \end{cases} \quad (1)$$

In equation (1), n represents the number of data. Based on M_3 and M_2 , the upper and lower edges of the abnormal data are obtained, as shown in equation (2).

$$\begin{cases} F_u = M_3 + \frac{1}{2} I_{MR} \\ F_l = M_1 - \frac{1}{2} I_{MR} \\ I_{MR} = M_3 - M_2 \end{cases} \quad (2)$$

In equation (2), I_{MR} represents the distance between M_3 and M_2 ; F_l represents the lower limit; F_u represents the upper limit. Therefore, data greater than F_u and less than F_l are considered abnormal data. After the abnormal data is diagnosed, CS-RBF is introduced in this study to correct the abnormal data. In radial basis function (RBF), the number of nodes in the input layer is typically decided by the dimensionality of the data set being imported [14]. The hidden layer usually converts unclassifiable data to higher dimensions based on kernel functions for classification, and the RBF activation function is shown in equation (3).

$$\varphi_j(x^i) = \exp\left(-\frac{1}{2\sigma^2} \|x^i - c_j\|^2\right) \quad (3)$$

In equation (3), c_j represents the center of the radial basis function; σ represents the width of the field; $\|x^i - c_j\|$ represents the distance between the two. The output of the output layer is mainly generated by the second layer through linear combination, so the output of RBF is shown in equation (4).

$$y(x^i) = w_0 + \sum_{j=1}^m w_j \varphi_j(x^i) \quad (4)$$

In equation (4), w_0 represents the offset; w_j represents the weight value between the output result of the j -th neuron in the second layer and the node; $\varphi_j(x^i)$ represents the output value of the activation function. CS is an algorithm derived from the cuckoo population with few parameters and excellent global search ability. This study optimizes RBF based on this algorithm. CS uses Levy flight to avoid getting stuck in local optima during the optimization process, and the search process is shown in equation (5) [15].

$$X_i^{t+1} = X_i^t + \alpha \otimes Le(\beta) \quad (5)$$

In equation (5), X_i^{t+1} represents the position of the i -th point at $t+1$; α represents step size coefficient; $\otimes$ represents dot multiplication operation method; $Le(\beta)$ represents random route. The Levy equation is shown in equation (6).

$$Levy(\beta) \sim u = t^{-\beta}, (1 \leq \beta \leq 3) \quad (6)$$

In equation (6), β represents the empirical value with a value of 1.5. Based on equation (6), a random step size calculation equation that satisfies this type of feature is integrated, as shown in equation (7).

$$\begin{cases} S = \frac{\mu}{|v|^{\frac{1}{\beta}}} \\ \begin{cases} \mu \sim N(0, \sigma_\mu^2) \\ v \sim N(0, \sigma_v^2) \end{cases} \\ \sigma_v = 1 \\ \sigma_u = \left\{ \frac{\Gamma(1+\beta) \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left[\frac{1+\beta}{2}\right] \beta 2^{\frac{\beta-1}{2}}} \right\}^{\frac{1}{\beta}} \\ \Gamma(z) = \int_0^\infty \frac{t^{z-1}}{e^t} dt \end{cases} \quad (7)$$

After obtaining the relevant calculation equation for the step size, when there are abandoned bird eggs, a new residence is searched for through local random flight, as shown in equation (8).

$$X_i^{t+1} = X_i^t + as \otimes H(Pa - r) \otimes (X_c^t - X_k^t) \quad (8)$$

In equation (8), X'_c and X'_k respectively represent the random two points at time t ; r is a random number with a value between $[0,1]$; H is the transition function; P_a represents the probability of being discovered. Based on traditional CS, this study optimizes CS from the perspectives of discovery probability and step length. There are mainly two situations to consider in terms of the probability of being discovered. On the one hand, if the probability of being discovered is too high, it may lead to a higher probability of the optimal solution being discarded. On the other hand, if the probability is low, poor solutions may not be processed in a timely manner, thereby affecting efficiency [15-16]. This study proposes relevant optimization strategies based on this, and the optimized discovery probability is shown in equation (9).

$$P_t = p_{\min} + (p_{\max} - p_{\min}) \left(1 - \frac{f_{\text{best}}^t}{f_i^t}\right) \quad (9)$$

In equation (9), t represents the population algebra; p_t represents the probability of an

individual being discovered and generating a new solution; f_i^t and f_{best}^t represent the function solutions corresponding to the i -th residence and the optimal residence, respectively. In view of the fact that the selection of step size in traditional CS often depends on human subjectivity, which leads to low computational efficiency of the algorithm, this study optimizes the location of the optimal residence, as shown in equation (10).

$$\begin{cases} d_i = \frac{\|n_i - n_{\text{best}}\|}{d_{\max}} \\ \text{step}_i = \text{step}_{\min} + (\text{step}_{\max} - \text{step}_{\min})d_i \end{cases} \quad (10)$$

In equation (10), n_i denotes the position of the i -th residence; $d_{\max}$ denotes the maximum distance between the current optimal residence and other residences; $\text{step}_{\min}$ and $\text{step}_{\max}$ represent the minimum and maximum step sizes, respectively. After obtaining the optimized CS, it is used to optimize RBF, and the neural network algorithm distance is shown in Figure 3.

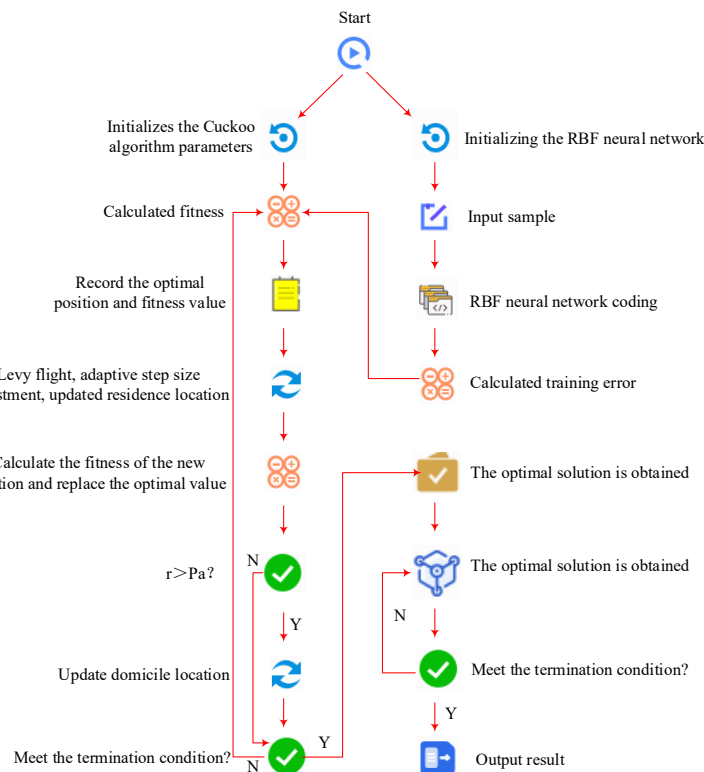


Figure 3: CS-RBF Network Flowchart

Figure 3 shows the specific process of CS-RBF, where CS optimizes RBF in the following way: the location information of each residence is converted into output layer weights in the network model, and based on CS, field width, and center position information, the place with the highest fitness value is identified to obtain the optimal parameters of RBF.

After the CS-RBF neural network corrects the outliers, the weighted average summation stage is performed. After the experimental environment is divided into several small areas, sensors are used to collect data from each target area and finally perform weighted summation; Each sensor is assigned corresponding weights to evaluate

differences in the environment. This study introduces an adaptive weighted average methodology for the fusion of the corresponding data from each sensor, and its value is shown in equation (11).

$$\begin{cases} X = \sum_{i=1}^n w_i x_i \\ \sum_{i=1}^n w_i = 1 \\ \sigma^2 = E[\sum_{i=1}^n w_i^2 (A - x_i)^2] = \sum_{i=1}^n w_i^2 \sigma_i^2 \end{cases} \quad (11)$$

In equation (11), n denotes the number of sensors; x_i denotes the measurement value of the i -th sensor; w_i denotes the weight corresponding to the i -th sensor; A denotes the estimated true value. The sensor data is arranged in ascending order, and the correlation distance is shown in equation (12).

$$\begin{cases} dM = M_h - M_l \\ g_1 = M_l - \frac{1}{2}dM \\ g_2 = Q_h + \frac{1}{2}dM \end{cases} \quad (12)$$

In equation (12), M_h represents the upper quartile; M_l represents the lower quartile; dM represents the distance between M_h and M_l ; interval $[g_1, g_2]$ represents the value of dM from the median.

2.2 Construction of Global Fusion Center Model Based on D-S

After weighted fusion of sensor data through local fusion centers, this study proposes a global fusion center based on D-S. The D-S evidence theory identification framework consists of a set. The subsets of this set all obtain a value between $[0,1]$, and the elements in the set are combined to form the corresponding power set. This power set has a mapping relationship with interval $[0,1]$, defined as the Basic Probability Assignment (BPA), as shown in equation (13).

$$\begin{cases} m(\phi) = 0 \\ \sum_{\theta \in 2^u} m(\theta) = 1 \\ 0 \leq m(\theta) \leq 1 \end{cases} \quad (13)$$

In equation (13), u represents the recognition framework; θ represents a subset of u ; 2^u represents a power set; $m()$ represents the probability of the corresponding subset. Based on the above mapping relationship, the trust function is shown in equation (14).

$$\begin{cases} Bel(A) = \sum_{B \in A} m(B) \\ Bel(\phi) = 0 \\ Bel(u) = 1 \end{cases} \quad (14)$$

In equation (14), $Bel()$ represents the trust function. Based on this mapping relationship, the likelihood function can be obtained, as shown in equation (15).

$$Pl(A) = 1 - Bel(A) = \sum_{B \cap A \neq \phi} m(B) \quad (15)$$

In equation (15), $Pl(A)$ denotes the likelihood function. The traditional D-S theory may suffer from poor stability of fusion results and the problem of veto power due to conflict coefficients [17]. This study improves the D-S theory to enable it to be used in conflicting data, introducing cosine angles to detect corresponding evidence conflicts, and obtaining the average similarity of evidence based on the generated matrix, as shown in equation (16).

$$\begin{cases} V_{ij} = \cos m_i m_j = \frac{(m_i, m_j)}{|m_i| \cdot |m_j|} \\ \bar{V} = \begin{bmatrix} \bar{V}_1 \\ \bar{V}_2 \\ \vdots \\ \bar{V}_n \end{bmatrix} \\ \bar{V}_1 = \frac{\sum_{j=1}^k V_{ij}}{k} \end{cases} \quad (16)$$

In equation (16), m_i and m_j represent independent evidence. Based on the size of the cosine value, the relationship between evidence is determined. If the function value is 1, the evidence is the same; When the function value is 0, there is no correlation between the evidence; When the function value is -1, it indicates a significant conflict between the evidence. The corresponding threshold is set, and after comparing with the average similarity obtained above, whether there is a conflict can be determined. Based on the relevant results of the cosine function mentioned above, if it is determined that there is a conflict between the evidence, the distance function is introduced to obtain a new basic probability allocation. Based on each piece of evidence, its distance from the mean is calculated as shown in equation (17).

$$\begin{cases} d_i = e^{-|m_{i1} - \bar{m}_1|} + e^{-|m_{i2} - \bar{m}_2|} + \dots + e^{-|m_{ij} - \bar{m}_j|} \\ \omega_i = \frac{d_i}{\sum_{i=1}^n d_i} \\ \bar{m}^{\sim} = \omega_i m_i \end{cases} \quad (17)$$

In equation (17), d_i denotes the distance between the i -th evidence and the mean; $\bar{m}_e$ represents the average value of all evidence based on the e -th focal element; m_{ij} represents the BPA function value of the evidence relative to the j -th focal element; ω_i represents the weight. Based on the weights corresponding to each evidence, first construct a reliability matrix Q , as shown in equation (18).

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1m} \\ m_{21} & m_{22} & \cdots & m_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nm} \end{bmatrix} \quad (18)$$

Based on equation (18), the transformed matrix can be obtained by performing relevant transposes and transformations on the matrix, as shown in equation (19).

$$A = \begin{bmatrix} m_{i1} \times m_{j1} & m_{i2} \times m_{j2} & \cdots & m_{im} \times m_{jm} \\ m_{i1} \times m_{j1} & m_{i2} \times m_{j2} & \cdots & m_{im} \times m_{jm} \\ \vdots & \vdots & \cdots & \vdots \\ m_{im} \times m_{j1} & m_{im} \times m_{j2} & \cdots & m_{im} \times m_{jm} \end{bmatrix} \quad (19)$$

Based on equation (19), several perfusion factors are fused as shown in equation (20).

$$\begin{cases} m_j = \frac{M_{jj}}{1 - K^-} \\ K^- = \sum_{i \neq j} m_{ii} \times m_{ji} \end{cases} \quad (20)$$

In equation (20), K^- represents the uncertainty coefficient. This study introduces the membership degree in fuzzy sets to evaluate the probability of a single target belonging to a certain interval. The domain is defined as U , where A is a fuzzy subset, so the relevant mapping relationship and membership function are shown in equation (21).

$$\begin{cases} \mu_A(n) : U \rightarrow [0,1], n \rightarrow \mu_A(n) \in [0,1] \\ \mu_A(x) = \begin{cases} L(x), l \leq c \leq m \\ R(x), m \leq c \leq r \end{cases} \end{cases} \quad (21)$$

In equation (21), μ_A represents the membership function; n represents an element; $L(x)$ and $R(x)$ are within the range of $[0,1]$. Furthermore, a membership function is introduced to evaluate the relevant probabilities, and the membership function corresponding to the basic probability allocation function is established, as shown in equation (22).

$$\begin{cases} \mu(A) = \begin{cases} (x - t_a) / (t_b - t_a) & t_a < x \leq t_b \\ (t_c - x) / (t_c - t_b) & t_b < x \leq t_c \\ 0 & x \leq t_a \text{ or } x > t_c \end{cases} \\ \mu(B) = \begin{cases} 1 & x \leq t_e \\ (t_e - x) / (t_e - t_d) & t_d < x \leq t_e \\ 0 & x > t_e \end{cases} \\ \mu(C) = \begin{cases} (x - t_f) / (t_g - t_f) & t_f < x \leq t_g \\ (t_h - x) / (t_h - t_g) & t_g < x \leq t_h \\ 0 & x \leq t_f \text{ or } x > t_h \end{cases} \end{cases} \quad (22)$$

In equation (22), $\mu()$ represents the membership function, and A , B , and C respectively represent temperature, humidity, and light intensity. Based on equation (22), obtain the basic probability distribution function of A , B , C , as shown in equation (23).

$$\begin{cases} m_i(H_j) = \frac{\mu_i(H_j)}{\sum_j \mu_i(H_j) + (1 - \alpha_i)(1 - \beta_i)} \\ m_i(\square) = 1 - m_i(H_1) - m_i(H_2) - m_i(H_3) \\ \alpha_i = \mu_i(H_k) - \max\{\mu_i(H_j)\} \quad j \neq k \\ \beta_i = \sqrt{\frac{1}{k-1} \sum_{j \neq k} [\mu_i(H_j) - w_i]^2} \\ w_i = \sum_{j \neq k} \frac{1}{k-1} \mu_i(H_j) \end{cases} \quad (23)$$

In equation (23), $m_i(H_i)$ denotes the basic probability allocation function; $m_i(\square)$ represents uncertainty description; α_i represents the difference between the maximum membership degree and the minimum membership degree; β_i represents the membership variance of the evidence after removing the maximum value. For the final fusion result, the following rules are set to determine the state information of the fusion result, as shown in equation (24).

$$\begin{cases} m(H_m) = \max\{m(H_i)\} & m(H_m) - m(H_j) > \eta \\ m(H_m) - m(\mathcal{G}) > \eta & \eta \in R, \eta > 0 \\ m(\mathcal{G}) < \lambda & \lambda \in R, \lambda > 0 \end{cases} \quad (24)$$

In equation (24), λ and η represent parameters greater than 0. If the final fusion result satisfies all the conditions in equation (24) simultaneously, it is determined that H_m has occurred in the hypothesis space. The process of the global fusion center is shown in Figure 4.

In Figure 4, the specific steps of the global fusion center are as follows: first, the local fusion center obtains the local fusion result through weighted

averaging, and then the BPA allocation function is constructed to obtain the original data source; Determine whether there are conflicts with this initial data source; If there is intrusion, evidence sources are fused with the improved D-S mentioned

above. Otherwise, evidence sources are fused based on general rules. If there are corresponding conflicts, the corresponding matrix also needs to be constructed to incorporate joint evidence factors.

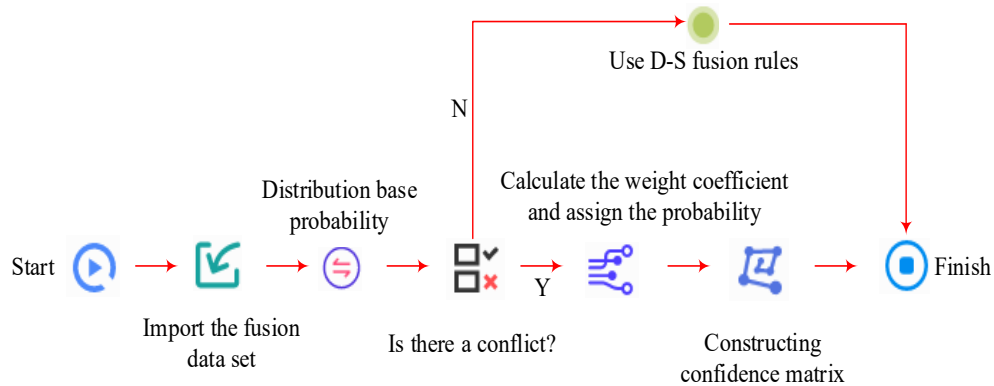


Figure 4: Global fusion algorithm flow

3. Results

3.1 Performance Evaluation Based on Improved CS-RBF Algorithm

This study selected the ripening process of cucumbers as relevant data, and collected data using three types of sensors including temperature. A total of 3300 sets of collected data were selected, with a training and testing data ratio of 10:1. The iteration times of the three algorithms were defined to stop at 250 times, and comparative experiments were conducted on the improved CS-RBF algorithm, the initial CS-RBF algorithm, and the PSO-RBF algorithm used in reference [18]. The fitness curves of the three algorithms are shown in Figure 5.

In Figure 5 (a), the improved CS-RBF tended to converge at 10 iterations, with the fitness value

converging to 110 and the initial fitness value being the smallest; The original CS-RBF showed a significant decrease before 50 iterations, and then fluctuated up and down with increasing iterations; Compared with the other two algorithms, PSO-RBF had a higher initial fitness value and a more significant decrease before the iteration number reached 30. The final convergence of the algorithm was similar to that of the original CS-RBF. In Figure 5 (b), overall, the recall rates of the three algorithms increased with the increase of iteration times. The improved CS-RBF recall rate ultimately converged to around 0.9, significantly higher than the other two algorithms. The predictive performance of three algorithms based on temperature, humidity, and light was further compared. The results are shown in Figure 6.

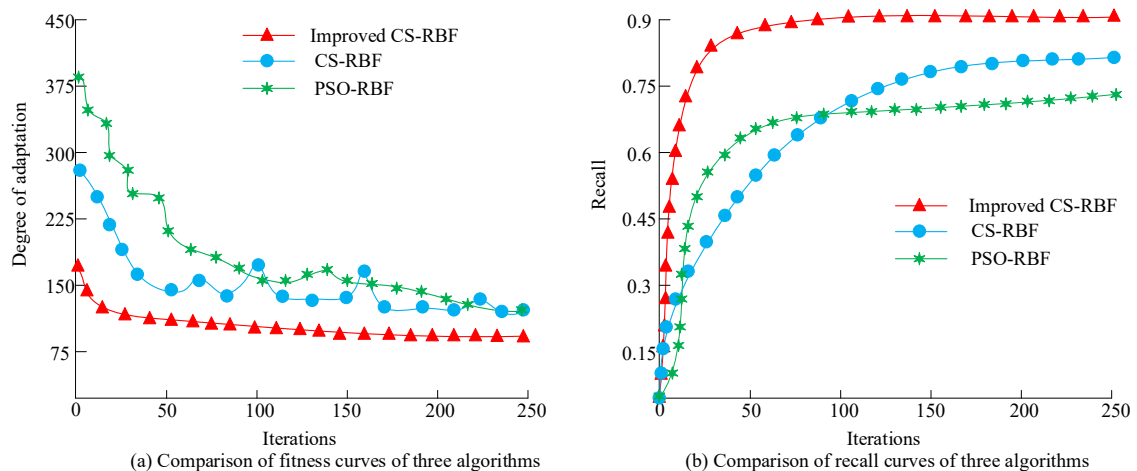


Figure 5: Comparison of three algorithm fitness and recall curves

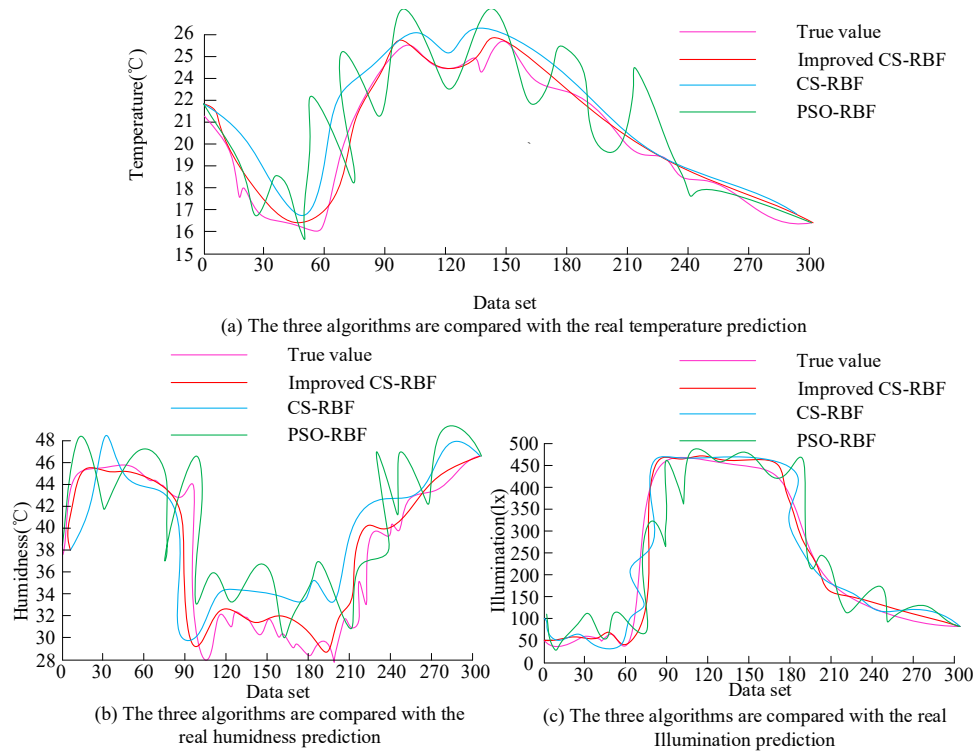


Figure 6: Comparison of three algorithms and actual values for temperature, humidity, and light prediction

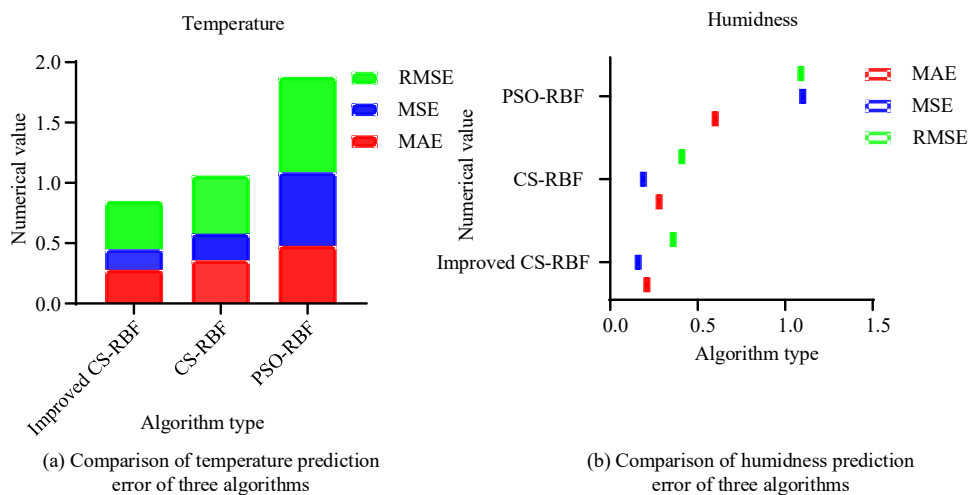


Figure 7: Comparison of temperature and humidity prediction errors of three algorithms

In Figure 6 (a), the improved CS-RBF was closest to the true value curve, followed by the original CS-RBF; The PSO-RBF algorithm was prone to getting stuck in local optima and had significant differences from the true value curve. In Figure 6 (b), the improved CS-RBF algorithm was closer to the true value curve than the other two algorithms, indicating that it had good prediction accuracy. In Figure 6 (c), similar to temperature and humidity, the improved CS-RBF exhibited good prediction accuracy, and its curve could well coincide with the true value curve. The performance indicators of the three algorithms were further compared, and the results are shown in Figure 7.

In Figure 7 (a), the improved CS-RBF was at the minimum value level among the three error values, with Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) of 0.28, 0.36, and 0.48, respectively. The error values were reduced by an average of 0.14, 0.25, and 0.24. In Figure 7 (b), the PSO-RBF error value was much higher than the other two algorithms, indicating that its prediction performance for humidity was poor. The optimized CS-RBF had the smallest error values of 0.21, 0.16, and 0.36, respectively, which were reduced by an average of 0.25, 0.49, and 0.39 compared to the other two algorithms. Three algorithms based on data prediction of lighting were further compared. The results are shown in Figure 8.

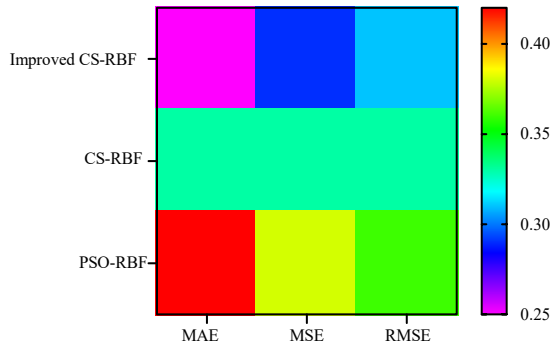


Figure 8: Comparison of illumination prediction error of three algorithms

In Figure 8, as the MSE value in the lighting data was one order of magnitude different from the other two error values, normalization method was used to visually evaluate the prediction error comparison of the three errors in the algorithms. In Figure 8, the improved CS-RBF was at the minimum numerical level in all three error values, while the PSO-RBF

error value was relatively large, far greater than the other two CS-RBF values. The improved CS-RBF algorithm had error values of 6124 and 8, respectively, which were reduced by an average of 3, 27, and 2 compared to the other two algorithms. Overall, the improved CS-RBF achieved the minimum error values in temperature, humidity, and lighting, indicating that CS-RBF simultaneously considered both global and local optimization capabilities.

3.2 Analysis of Global Fusion Results Based on D-S

Based on the local fusion experimental results obtained above, a set of typical data was characterized for the global fusion effect. The temperature was set to 21 °C, the humidity was 54%, and the illumination was 240lx. The fuzzy membership degrees and allocation functions of temperature, humidity, and lighting were constructed. Table 1 presents the results.

Table 1. Fuzzy membership degree and distribution function of temperature, humidity and illumination

	A	B	C	m(A)	m(B)	m(C)	m($\odot$)
Temperature	0	0.398	0.612	0	0.241	0.384	0.394
Humidness	0	0.336	0.686	0	0.221	0.445	0.361
Illumination	0	0.246	0.776	0	0.172	0.501	0.301

In Table 1, after obtaining the basic trust allocation function, the corresponding similarity matrix and average similarity matrix were obtained by performing corresponding matrix changes. This study selected an angle of 60 ° as the threshold and a corresponding cosine value of 0.5. The matrix results are shown in Table 2.

Table 2. Similarity and average similarity matrix

Matrix name	Base vector		
Similarity	1	0.996	0.962
	0.996	1	0.982
	0.962	0.982	1
Average similarity	0.986	/	/
	0.992	/	/
	0.981	/	/

In Table 2, the similarity between all evidence was greater than the threshold of 0.5 and exhibited a high level of numerical hierarchy, indicating that the conflict was acceptable. Based on the D-S combination rule, relevant evidence was combined for research. The results are shown in Figure 9.

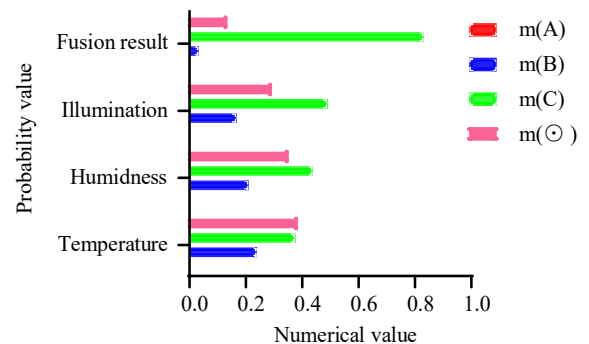


Figure 9: Result graph based on D-S evidence combination rule fusion

In Figure 9, the credibility of the C focal element was higher than that of the other focal elements, with a fusion result value of 0.832. Based on the specific values of the original data, measures such as supplementing light intensity and using fans to increase humidity should be taken at this time. The results of this data indicated that the fusion effect met the expected goals. Further research selected a new set of data, namely temperature of 23 °C, light intensity of 250lx, and humidity of 66%. The relevant fusion results are shown in Figure 10.

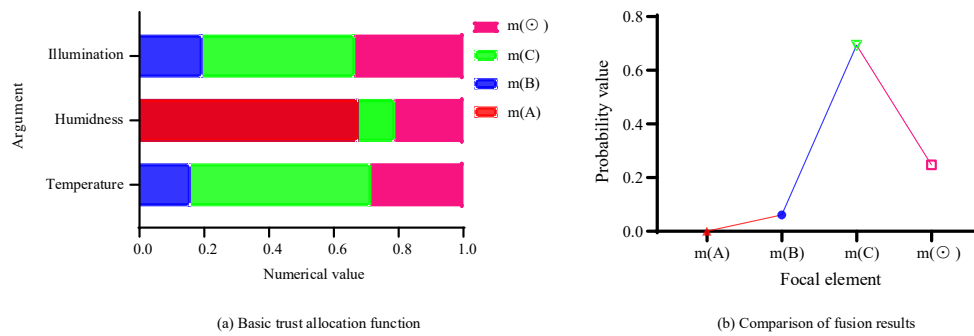


Figure 10: Basic trust allocation and comparison of fusion results

In Figure 10 (a), the probability value of temperature and light based on focal element C was the highest, while the probability value of humidity based on focal element A was 0.677. The trust allocation function established based on this had specific cases where the average similarity was less than the threshold. Therefore, the improved D-S theory was introduced to fuse this system, and the fusion result was shown in Figure 6 (b). At this point, the probability of focusing on element C was still the highest, which means that in this case, it was necessary to supplement the light intensity and use fans to increase humidity. By comparing Figure 10 (a) and Figure 10 (b), the uncertainty based on the final result decreased from 32.3% to 23.9%, while the output probability increased from 11.3% to 68.5%. Thus, the enhanced D-S could significantly reduce uncertainty, which could also handle the issue of veto power.

4. Discussion

In the experimental results, the improved CS-RBF tended to converge at 10 iterations, and the fitness value converged to 110. Meanwhile, the improved CS-RBF recall rate eventually converged to around 0.9. The improved CS-RBF was closest to the true value curve, and it was at the minimum level among the three error values. Its MAE, MSE, and RMSE are reduced by an average of 0.14, 0.25, and 0.24 compared to the other two algorithms. The humidity was reduced by an average of 0.25, 0.49, and 0.39, and the lighting was reduced by an average of 3, 27, and 2. In the first example, the C-element value was 0.832. In contrast, in the second example, the introduction of the improved D-S theory reduced the uncertainty of the final result from 32.3% to 23.9%, while increasing the output probability from 11.3% to 68.5%. This indicated that the improved D-S theory could significantly reduce uncertainty and solve the problem of veto power. Currently, scholars have supported the viewpoints in the research. Dai et al. optimized the Elman network with improved CS to improve the accuracy of the electric spindle error model, indicating that the network model constructed in this study had better prediction accuracy and stability, while significantly reducing

the error value. This supported the performance results of the enhanced CS in the research [19]. To study the complex evolution process of a major rainstorm disaster in China, Xie et al. proposed a Bayesian network combined with D-S theory to explore the impact of meteorological factors, emergency activities and other factors on the disaster evolution mechanism. The findings indicated that the enhanced D-S evidence theory has the potential to mitigate the subjectivity inherent in the model's approach to uncertainty, quantify and reduce the impact of human emotion changes, and meanwhile, better replicate the actual events, which was consistent with the improved D-S theory in reducing uncertainty [20].

In summary, the improved CS-RBF proposed in the study not only has global search capability but also excellent local search capability. Meanwhile, the prediction errors for temperature, humidity, and light indicators are relatively small. The improved D-S theory can significantly reduce uncertainty and avoid the problem of veto power.

5. Conclusions

In the research on greenhouse environment detection experiments, an optimized CS-RBF is proposed. In the correction of outliers, the optimized CS-RBF algorithm is introduced, which optimizes CS in terms of discovery probability and step size based on the initial algorithm. The relevant experimental results demonstrated that the improved CS-RBF calculation efficiency, recall rate, and related error values were superior to the initial CS-RBF and PSO-RBF. At the level of global fusion center research, an improved D-S theory was introduced to fuse the data collected by local fusion centers. The results showed that the improved D-S theory could effectively avoid the veto problem and reduce uncertainty. Although the improved CS-RBF and D-S theories proposed in the study have superior data prediction for temperature, humidity, and light in greenhouse environments, there are also certain limitations in the research. Environmental factors in greenhouse environments are interrelated and influence each other, and other environmental factors will also change accordingly. In future

research, additional environmental factors will be added on top of the three factors to jointly study the performance and prediction effectiveness of the improved algorithm.

Fundings

The research is supported by Fujian Province Young and Middle-aged Teacher Education Research, Research on Application of Radar Sensors in IoT Systems, Project under Grant JAT210838.

Reference

- [1] Chen, M., Cai, Z., Zeng, Y., & Yu, Y. Z. (2023). Multi-sensor data fusion technology for the early landslide warning system. *Journal of Ambient Intelligence and Humanized Computing*, 14(8), 11165–11172.
- [2] Preethi, P., & Mamatha, H. R. (2023). Region-based convolutional neural network for segmenting text in epigraphical images. *Artificial Intelligence and Applications*, 1(2), 119–127.
- [3] Osornio-Rios, R. A., Zamudio-Ramírez, I., Jaen-Cuellar, A. Y., et al. (2023). Data fusion system for electric motors condition monitoring: An innovative solution. *IEEE Industrial Electronics Magazine*, 17(4), 4–16.
- [4] Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., et al. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, 96(1), 156–191.
- [5] Cañadas-Aránega, F., Blanco-Claraco, J. L., Moreno, J. C., & Moreno, J. C. (2024). Multimodal mobile robotic dataset for a typical Mediterranean greenhouse: The GREENBOT dataset. *Sensors*, 24(6), 18740.
- [6] Traça, A. (2023). Deep learning for autonomous vehicle sensor data analysis and interpretation. *Journal of Artificial Intelligence Research and Applications*, 3(1), 167–188.
- [7] Chen, J., Zhou, Z., Kim, B. J., Kim, B. J., Zhou, Y., Wang, Z. Q., et al. (2023). Optoelectronic graded neurons for bioinspired in-sensor motion perception. *Nature Nanotechnology*, 18(8), 882–888.
- [8] Yin, T., Xu, L., & Gil, B. (2023). Graphene sensor arrays for rapid and accurate detection of pancreatic cancer exosomes in patients' blood plasma samples. *ACS Nano*, 17(15), 14619–14631.
- [9] Fei, S., Hassan, M. A., Xiao, Y., Su, X., Chen, Z., Cheng, Q., et al. (2023). UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Precision Agriculture*, 24(1), 187–212.
- [10] Senel, N., Kefferpütz, K., Doycheva, K., Doycheva, K., & Elger, G. (2023). Multi-sensor data fusion for real-time multi-object tracking. *Processes*, 11(2), 501.
- [11] Liu, W., Liu, Y., & Bucknall, R. (2023). Filtering-based multi-sensor data fusion algorithm for a reliable unmanned surface vehicle navigation. *Journal of Marine Engineering & Technology*, 22(2), 67–83.
- [12] Cao, T. (2023). Effective detection algorithm of electronic information and signal processing based on multi-sensor data fusion. *The Egyptian Journal of Remote Sensing and Space Science*, 26(3), 519–526.
- [13] Dey, K., Kalita, K., & Chakraborty, S. (2023). Prediction performance analysis of neural network models for an electrical discharge turning process. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 17(2), 827–845.
- [14] Abdullah, H., & Noori, N. A. (2024). Comparison of non-linear time series models (Beta-t-EGARCH and NARMAX models) with radial basis function neural network using real data. *Iraqi Journal for Computer Science and Mathematics*, 5(3), 26–44.
- [15] An, J., Li, H., Ng, D. W. K., & Yuen, C. (2023). Fundamental detection probability vs. achievable rate tradeoff in integrated sensing and communication systems. *IEEE Transactions on Wireless Communications*, 22(12), 9835–9853.
- [16] Fahim, P., & Meira-Góes, R. (2024). Detecting probability footprints of sensor deception attacks in supervisory control. *IFAC-PapersOnLine*, 58(1), 192–197.
- [17] Punetha, T., Komal, & Pamucar, D. (2024). Novel picture fuzzy power partitioned Hamy mean operators with Dempster-Shafer theory and their applications in MCDM. *Artificial Intelligence Review*, 57(6), 145–152.
- [18] Ningning, S., & Ying, W. (2024). Research on PSO-RBF traction transformer fault diagnosis based on Adam optimization. *Journal of Electrical Engineering*, 18(4), 209–216.
- [19] Dai, Y., Wang, X., Li, Z. L., He, S., Yu, B. L., & Zhou, X. W. (2024). Thermal error modeling of electric spindles based on cuckoo algorithm optimized Elman network. *The International Journal of Advanced Manufacturing Technology*, 132(3), 1365–1375.
- [20] Xie, X., Tian, Y., & Wei, G. (2023). Deduction of sudden rainstorm scenarios: Integrating decision makers' emotions, dynamic Bayesian network, and DS evidence theory. *Natural Hazards*, 116(3), 2935–2955.