Research on Temperature Prediction Model in Greenhouse Based on Improved SVR

Linan Fan*, Yandong Ji, Gang Wu
School of Information Engineering, Shenyang University, China

*Corresponding author: linanfan@syu.edu.cn

Abstract. The greenhouse is a multivariate, nonlinear, time-varying and time-lag system. The establishment of an accurate greenhouse temperature prediction model can effectively ensure the effectiveness of intelligent temperature control. In order to solve the problem of difficulty in modeling the temperature mechanism of greenhouses, this paper proposes a method based on improved Support Vector Regression to establish greenhouse temperature prediction. The PSO algorithm is used to optimize the hyper-parameters of the SVR. Using real greenhouse data to verify the effectiveness of the proposed model. Experimental results show that compared with the traditional BP neural network prediction model, the temperature prediction model in greenhouse based on improved SVR has significantly improved prediction accuracy and better prediction capabilities. Therefore, the greenhouse temperature prediction model based on improved SVR proposed in this paper can well characterize the actual greenhouse system.

1. Introduction
Greenhouse is a multivariate, nonlinear, time-varying and time-lag system. There are many factors that affect the temperature change of the greenhouse. There are not only the operating conditions of the various actuators in the greenhouse, but also the temperature, light, wind speed, wind direction and other factors outside the greenhouse, as well as the physiological effects of the plant itself. Therefore, it is very difficult to establish a model that can cover all factors and perfectly reflect the conditions of the greenhouse. The traditional mechanism modeling method wants to obtain a more accurate model. As the factors involved in the greenhouse increase, the model complexity is also increasing [1]. The method of identification modeling is solving multivariate, nonlinear, and time-varying Time lag and other aspects have a strong ability [2].

The temperature is control variable of the greenhouse. In order to achieve the purpose of temperature control, it is necessary to establish the relationship model between the variables of the greenhouse and the temperature, that is, the greenhouse temperature prediction model. The input of the prediction model includes opening degree of sunroof, size of heater, size of fan, shading net switch, heat preservation shutter switch, wet curtain switch, exhaust fan switch, fill light switch, indoor air humidity, soil temperature, and soil humidity., Carbon dioxide concentration, light intensity, outdoor temperature, outdoor wind speed, the output of the prediction model is temperature.
2. Support Vector Regression

Support Vector Regression [3] is an important application of support vector machines in the field of regression. It can minimize structural risks and solve specific problems such as nonlinearity and high dimensionality in the case of limited samples [4]. The structure of the SVR is shown in Figure 1.

![The structure of the SVR](image)

**Fig 1.** The structure of the SVR

The input layer is the input data, the hidden layer selects the appropriate kernel function $K(x_i,x)$, and the last layer is the construction regression function $f(x)$. The main problem that the SVR algorithm solves is that a given n-dimensional input vector of a k sample set $\{(x_1,y_1),(x_2,y_2)\ldots(x_i,y_i)\ldots(x_k,y_k)\}$ (in the formula $x_i \in R^n$, $y_i \in R$ is the corresponding output data. The sample set is mapped to a high-dimensional feature space $H$ through a nonlinear mapping function $\Phi(x)$, and the hyperplane is searched for Linear regression. The regression form can be expressed as:

$$y(x) = w^T \Phi(x) + b$$  \hspace{1cm} (1)

Among, $w$ is the support vector and $b$ is the offset.

Introducing the slack variable $\xi_i$ and $\xi^*$ can better solve the uncertainty, so the SVR algorithm can be expressed as:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$$
$$\text{s.t. } f(x_i) - y_i \leq \xi_i + \xi_i^*,$$
$$y_i - f(x_i) \leq \varepsilon + \xi_i^*,$$
$$\xi_i, \xi_i^* \geq 0, i = 1, 2, \ldots, l$$  \hspace{1cm} (2)

Among, $C$ is the penalty factor, $\varepsilon$ is the empirical error, $C$ controls the empirical risk error of the SVR model. The larger $C$, the greater the punishment for samples with greater errors $\varepsilon$ during the training process. Therefore, the $C$ choice is very critical.
The above formula shows that the SVR algorithm has constraints and is a convex quadratic programming problem. The dual problem can be solved by using the Lagrange multiplier method, so the above formula can be expressed as:

\[
\begin{align*}
\min_{\alpha, \alpha^*} & \quad \sum_{i=1}^{l} y_i (a_i^* - a_i) + \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j^* (a_i^* - a_i) (a_j^* - a_j) \kappa(x_i, x_j) \\
\text{s.t.} & \quad \sum_{i=1}^{l} \alpha_i = \sum_{i=1}^{l} a_i^* = 0 \\
& \quad 0 \leq a_i, a_i^* \leq C
\end{align*}
\]

(3)

In this paper, the greenhouse temperature prediction model, the kernel function selects the radial basis function \( \kappa(x, x_i) = \exp(-\gamma|x - x_i|^2) \), which comes with a parameter \( \gamma \) that represents the bandwidth of the RBF. As \( \gamma \) increases, the accuracy of the training set is improved, which is likely to cause overfitting and lack generalization. As \( \gamma \) decreases, the accuracy of the training set is reduced, and it is easy to cause underfitting, and an ideal model cannot be obtained. Therefore, the choice of \( \gamma \) is related to whether the final model can truly represent the system.

Through the SVR mathematical formula, the equation (3) can be solved to obtain:

\[
a = [a_1, a_2, \ldots, a_l], \quad a^* = [a_1^*, a_2^*, \ldots, a_l^*].
\]

The regression function is:

\[
f(x) = \sum_{i=1}^{l} (a_i - a_i^*) \kappa(x, x_i) + b^*
\]

(4)

The SVR algorithm can be expressed in the form of the SVR function. Through the linear combination of the intermediate nodes, the support vector corresponding to each node is finally solved to obtain the regression function model.

The greenhouse temperature prediction model is essentially regression. Whether it can truly reflect the actual situation of the greenhouse has a very large relationship with the feature selection. The greenhouse temperature is not only related to the outside temperature and wind speed, but also the air humidity inside the greenhouse and the operating status of the actuator. Therefore, it is necessary to analyze the influence factors of the greenhouse temperature prediction model from different dimensions.

In order to better find the correlation between related variables and temperature in greenhouses, this paper simplified the input variables of the greenhouse temperature prediction model based on principal component analysis, and used several important principal components to replace all components to establish the greenhouse temperature prediction model. This paper establishes a greenhouse prediction model with greenhouse-related data as influencing factors, namely, skylights, heaters, fans, sunshade nets, thermal insulation curtains, wet curtains, exhaust fans, supplementary lights, air humidity, soil temperature, and soil humidity. Carbon dioxide concentration, light intensity, outdoor temperature, outdoor wind speed. The 15 impact factors were analyzed by principal components, and the specific variance and cumulative variance contribution rate are shown in Table 1. You can see the variance and class variance contribution rate of the 15 impact factors. The first 7 impact factors contain 99.19% of the total information, so the first 7 principal components are selected as the input factors of the model.

In order to ensure that the prediction model contains all the important features that affect the temperature of the greenhouse, and reduce the expression of the greenhouse because of fewer features, based on the characteristics of the principal component analysis after dimensionality reduction, The temperature prediction model based on SVR greenhouse contains 7 factors.
3. Prediction model of greenhouse temperature based on improved SVR

Through the basic introduction to the SVR model, we can see that there are two very important parameters for the SVR model that uses the RBF radial basis function, one is the penalty factor $C$, and the other is the width of the RBF kernel function $\gamma$. The selection of $C$ and $\gamma$ directly affects the regression model accuracy and generalization ability of the SVR model\cite{5,6}. In this paper, the SVR temperature prediction model parameter optimization uses the PSO algorithm, and the final calculated optimal solution is 17.325, and the optimal solution is 0.01.

When the SVR algorithm establishes a regression model for time series data such as greenhouse temperature, problems such as time lag and poor prediction accuracy often occur. Based on the greenhouse temperature in the actual operation process, more redundant data will be generated, so the fully connected form between the input layer and the middle layer is converted to a random connection form, and at the same time, due to the start and stop history of the main actuators The impact on the temperature of the greenhouse, so this paper uses the start and stop status of the historical actuator as an additional input to the SVR to ensure the retention of the main time series data characteristics.

| Code | Factors                        | Contribution (%) | Cumulative contribution (%) |
|------|--------------------------------|------------------|------------------------------|
| S1   | outdoor temperature           | 48.11%           | 48.11%                       |
| S2   | light intensity               | 27.63%           | 75.74%                       |
| S3   | heater                        | 11.91%           | 87.65%                       |
| S4   | heat preservation shutter     | 9.15%            | 96.80%                       |
| S5   | soil temperature              | 1.16%            | 97.96%                       |
| S6   | sunroof                       | 0.89%            | 98.85%                       |
| S7   | wind speed                    | 0.34%            | 99.19%                       |
| S8   | exhaust fan                  | 0.26%            | 99.45%                       |
| S9   | air humidity                  | 0.17%            | 99.62%                       |
| S10  | wet curtain                   | 0.08%            | 99.70%                       |
| S11  | shading net                   | 0.08%            | 99.78%                       |
| S12  | soil humidity                 | 0.07%            | 99.85%                       |
| S13  | carbon dioxide concentration  | 0.07%            | 99.92%                       |
| S14  | fill light                    | 0.07%            | 99.99%                       |
| S15  | fan                           | 0.01%            | 100%                         |

4. Example verification

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.1. Data sample selection and data normalization processing

The data used in this paper is the spring data of Northeast greenhouses, the crops are tomatoes, and the collection time is between 0:00 on March 31 and 24:00 on April 4. The data collection interval is 10 minutes, and the data includes sunroof opening, heater size, fan size, shading net switch, heat preservation shutter switch, wet curtain switch, exhaust fan switch, fill light switch, air humidity, soil temperature, soil humidity, carbon dioxide concentration, light intensity, outdoor temperature, outdoor wind speed, a total of 15 characteristics, the temperature in the greenhouse is used as the model output, the principal component analysis above uses the outdoor temperature, light intensity, heater size,
The insulation shutter switch, soil temperature, 7 sunroof opening and wind speed are used as model input features. A total of 720 groups of collected data. The specific data format is shown in Table 2.

**Table 2. Form of greenhouse data**

| No. | sunroof | heat preservation shutter | heat | wind speed/m/s | soil temperature/oC | light intensity/04Lx | outdoor temperature/oC | indoor temperature/oC |
|-----|---------|--------------------------|------|---------------|---------------------|--------------------|------------------------|------------------------|
| 1   | 0       | 1                        | 0.55 | 3             | 15.3                | 0                  | -2.4                   | 16.7                   |
| 2   | 0       | 1                        | 0.55 | 7             | 15.2                | 0.02               | -2.5                   | 16.7                   |
| 3   | 0       | 1                        | 0.55 | 7             | 15.3                | 0.02               | -2.7                   | 16.9                   |
| 4   | 0       | 1                        | 0.55 | 8             | 15.2                | 0.04               | -2.9                   | 16.8                   |
| 5   | 0       | 1                        | 0.55 | 8             | 15.2                | 0.02               | -3                     | 16.9                   |

......

The opening degree of the sunroof is divided into closed, 1/3 open, 2/3 open, and fully open, which are represented by 0, 1/3, 2/3, and 1, respectively. The heater 3kW is divided into nine levels, with 1-9 indicates that the thermal insulation shutter switch is represented by 1, 0.

Because there are many variables that affect the temperature of greenhouses, and these variables are in different quantities, there are obvious differences in the data. Therefore, before the establishment of the greenhouse temperature prediction model, the data is normalized. In this paper, the maximum normalization method [7] is used to linearly change the original data sample, and the data is uniformly mapped to the quantity [0,1]. The specific normalization formula is as follows:

\[
y_i = \frac{x_i - 0.95x_{\text{min}}}{1.05x_{\text{max}} - 0.95x_{\text{min}}}\]

Where \( y_i \) is the normalized value, \( x_i \) is the original value, \( x_{\text{min}} \) is the minimum value in the data, and \( x_{\text{max}} \) is the maximum value in the data.

4.2. Model accuracy evaluation standard

BP neural network is the most commonly used algorithm in greenhouse temperature intelligent control modeling [8-12], and it has good prediction accuracy. This paper constructs a BP neural network greenhouse temperature prediction model, and compares it with the improved SVR greenhouse temperature prediction model proposed in this paper. Figure 2 is the modeling result of the BP neural network model, and Figure 3 is the modeling result of the optimized SVR model. The circle in the figure represents the original temperature, and the star line represents the regression predicted temperature. It can be seen from Figures 2 and 3 that the BP neural network model can make a good prediction of the temperature trend, but it has a certain hysteresis in the prediction. The improved SVR model can well represent the actual temperature and predict the temperature. The accuracy is higher and the lag of prediction is reduced to a certain extent. The improved SVR model is closer to the actual result than the BP neural network model. The average relative error of the optimized SVR modeling result is small, the overall relative error is relatively stable, and the prediction result is within the allowable range of accuracy. The average relative error of the BP neural network modeling results is too large, the overall relative error fluctuates greatly, and some prediction results are outside the allowable range of accuracy.
Fig 2. The temperature prediction model of BPNN

Fig 3. The temperature prediction model of improve SVR

Table 3 shows the evaluation indicators of the two results. It can be seen from Table 3 that the maximum relative error of the optimized SVR model is 1.73%, and the maximum relative error of the BP neural network is 16.23%. Through comparison, it is found that the BP neural network has achieved good results in the temperature prediction of the greenhouse, and the optimized SVR model is significantly better than the BP neural network in modeling accuracy. The optimized SVR model proposed in this paper is an algorithm used for small samples such as greenhouses. After using PSO to optimize the two important parameters of the SVR model, the SVR model can be better constructed, and then the temperature of the greenhouse Intelligent control modeling.

| MODEL         | E      | MSE    | R2      |
|---------------|--------|--------|---------|
| Improved SVR  | ≤1.73% | 0.00019| 0.99945 |
| BPNN          | ≤16.23%| 0.019564| 0.96425 |
5. Conclusion
For the greenhouse temperature prediction, based on the principal component analysis, 7 factors were extracted as the input features of the model, and the greenhouse temperature prediction model based on the improved SVR algorithm was established, and the real greenhouse temperature data was used for prediction. The overall relative error is less than or equal to 1.73%, the root mean square error is 0.00019, and the coefficient of determination is 0.99945. Compared with the traditional BP neural network prediction model, the optimized SVR temperature prediction model has relatively stable prediction results, high prediction accuracy, and strong generalization ability. At the same time, for the lag of the prediction model of the BP neural network, the problem of the prediction lag of the prediction model is effectively improved by converting the fully connected form between the input layer and the intermediate layer into a random connection form. The greenhouse temperature prediction model establishes the relationship model between the variables of the greenhouse and the temperature. It provides the basis for the regulation of the intelligent temperature control system of the greenhouse.

Acknowledgments
This work was supported by LiaoNing Revitalization Talents Program under Grant (XLYC1806016).

References
[1] Zhang F, Liu D. Design of greenhouse data acquisition and control system[J]. Modernizing Agriculture,2018,17(04):2-4.
[2] Fan L N, Liu Z, Wu G, Li J Y, Dai X. Electronic Instrumentation Customers[J]. Electronic Instrumentation Customers,2019,26(01):6-9.
[3] Zhang C L, Liu J, Li X. Collaborative Prediction Method of Multi-Bearing Remaining Useful Life Based on the Improved PSO-SVR Model[J]. Machine Tool & Hydraulics,2020,48(16):206-211.
[4] Chen D Y, Chen D W, Chen K H, Xiao L D, Jiang S X. Study on subway passenger travel time prediction based on passenger flow big data analysis and support vector regression[J]. Modern Urban Transit,2020,3(09):70-76.
[5] Li S Q, Liu L L, Rong J, Zhou W, Liu L B. Atmospheric weighted mean temperature based on support vector regression[J]. Journal of Guilin University of Technology,2019,39(03):656-660.
[6] José Manuel Valente, Sebastián Maldonado. SVR-FFS: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression[J]. Expert Systems With Applications,2020,17(04):160-165.
[7] Neuroscience; A correspondence between normalization strategies in artificial and biological neural networks[J]. Computers Networks & Communications,2020,35(04):11-18.
[8] Liu S M, Xue Q Y, Li Z F, Li C, Gong Z H, Li N. BP neural network based solar greenhouse temperature prediction model[J]. Journal of China Agricultural University,2015,20(01):176-184.
[9] Li H, Tian F M, Tan F, Zhu P P. Journal of Agricultural Mechanization Research[J]. Journal of Agricultural Mechanization Research,2020,43(06):189-193.
[10] Wang B, Chen B W, Wang G Q, Li R, Wen J M, Lu C, Tian R F, Deng J. Back propagation (BP) neural network prediction and chaotic characteristics analysis of free falling liquid film fluctuation on corrugated plate wall[J]. Annals of Nuclear Energy,2020,39(07):148-151.
[11] Tian F M, Li H, Zhu P P. Research on North Experimental Greenhouse Temperature Prediction Model Based on Meteorological Data[J]. Journal of Heilongjiang Bayi Agricultural University,2020,32(04):73-80+113.
[12] Zhang K X, Liu Y Q, Guo J Y, Hu S Y, Wu Z G. Prediction Model of Primary Air Flow of Pulverizer Based on PCA and the GA-BP[J]. Electronic Instrumentation Customers,2017,24(04):15-18+100.