A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms

Jinglin Du [1,2,*], Yayun Liu [1,2], Yanan Yu [1,2] and Weilan Yan [1,2]

[1] School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20152281493@nuist.edu.cn (Y.L.); 20152281517@nuist.edu.cn (Y.Y.); 20151242455@nuist.edu.cn (W.Y.)

[2] Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing 210044, China

* Correspondence: jldunuistgmu@gmail.com; Tel.: +86-25-5873-1196

Abstract: Precipitation is a very important topic in weather forecasts. Weather forecasts, especially precipitation prediction, poses complex tasks because they depend on various parameters to predict the dependent variables like temperature, humidity, wind speed and direction, which are changing from time to time and weather calculation varies with the geographical location along with its atmospheric variables. To improve the prediction accuracy of precipitation, this context proposes a prediction model for rainfall forecast based on Support Vector Machine with Particle Swarm Optimization (PSO-SVM) to replace the linear threshold used in traditional precipitation. Parameter selection has a critical impact on the predictive accuracy of SVM, and PSO is proposed to find the optimal parameters for SVM. The PSO-SVM algorithm was used for the training of a model by using the historical data for precipitation prediction, which can be useful information and used by people of all walks of life in making wise and intelligent decisions. The simulations demonstrate that prediction models indicate that the performance of the proposed algorithm has much better accuracy than the direct prediction model based on a set of experimental data if other things are equal. On the other hand, simulation results demonstrate the effectiveness and advantages of the SVM-PSO model used in machine learning and further promises the scope for improvement as more and more relevant attributes can be used in predicting the dependent variables.

Keywords: support vector machine; particle swarm optimization; precipitation prediction

1. Introduction

Weather prediction is always a challenging problem, and many weather forecasters and experts devote themselves to improving the accuracy of prediction. For example, weather research and forecasting models (WRF), automatic weather stations (AWS), and numerical weather prediction models (NWP) and so on are proposed and used in practice [1]. This study is an attempt to make use of the efficient data mining techniques. In recent years, data mining is widely used for many machine learning problems such as bio-informatics, pattern recognition, linear regression analysis and nonlinear regression estimation problems. It is applied in several areas and proven effective when solving those problems, which has provided procedures for both descriptive (characterizations of the properties of the data) and predictive (learning and induction of the data for forecasting) tasks [2–4]. The selection of the classifier has a crucial impact on the accuracy and efficiency of monitoring. Annas et al. [5] presented the neural-fuzzy classification model in order to assess the rainfall variations in a tropical

area. This model derives the cluster information for the datasets by using the fuzzy C-means (FCM) clustering. The proposed method can improve the interpretability of variation in rainfall, but it cannot provide an exact solution to the problem. In addition, it is unreliable and too complicated to achieve. Prasad et al. [6] proposed an efficient decision tree algorithm to predict precipitation based on the historical climate data, obtained from the meteorological department. This model is an easily implementable tool, but the accuracy is not high, and the efficiency is low when dealing with large amounts of data. Recently, due to Support Vector Machine's (SVM's) excellent characteristics of robustness and generalization performance, many researchers considered the SVM method in order to improve the accuracy of the rainfall prediction model. Lu and Wang [7] predicted monthly rainfall in a region of China by using the SVM approach with different kernel functions. Kisi and Cimen [8] mentioned a novel wavelet-SVM approach to precipitation forecasting from past data. SVMs have also been recently applied to precipitation related studies, such as: Ortiz-Garcia [9] considered a large set of novel predictive variables in the daily precipitation prediction problem and analyzed the importance of humidity and Equivalent Potential Temperature variables in the SVM approach. SVM showed excellent performance by comparison with K-nearest neighbor (KNN) and multi-layer perceptron (MLP) approaches. However, the novel predictive variables are not easy to get in daily life. Sanchez-Monedero [10] established a model of rainfall occurrence and amounts simultaneously by using support vector classifiers. In addition, they used a novel set of predictive meteorological variables that improve the classifier's performance with this problem. Sehad [11] proposed a novel SVM-based technique to improve rainfall estimation using the multispectral Meteosat Second Generation (MSG) and Spinning Enhanced Visible and Infra Red Imager (SEVIRI) imagery. Younga [12] proposed a new hybrid approach integrating the physically based (Hydrologic modeling system, HEC-HMS) and data-driven (Support vector regression, SVR) models to predict the hourly runoff discharges in the Chishan Creek basin, southern Taiwan. Take SVM for accurate rainfall-runoff modeling. After a lot of work using the SVM algorithm has been done on the precipitation forecast, there is still some open research for improving the performance of the systems. Above all, in this paper, we propose a prediction model for rainfall forecasts based on Support Vector Machine (SVM). The algorithm shows many unique advantages in solving small samples, nonlinear and high dimensional pattern recognitions. The learning problem of SVM can be expressed as a convex optimization problem, so we can find the global minimum of the objective function by using the known effective algorithm [13]. The SVM is very efficient since there only two parameters that need to be optimized. It also avoids the dimensionality of the disaster owing to the fact that the final model is only related to the support vector.

SVM is a data mining technique based on machine learning used for data classification. One may use a classification algorithm to predict whether the particular day will be sunny or rainy. SVM is one of the most advanced classification techniques [14]. Benchmarking studies have shown that the SVM performs the best among current classification techniques [15]. Numerous experiments have shown that Support Vector Machine has satisfactory classification accuracies under a limited number of training samples, and it has been widely used in classification and prediction [9–12,16,17]. However, there is still a problem that the proper selection of kernel function and its parameters has great influence on the final prediction accuracy. In order to choose the optimal parameters when building the model, some researchers have linked Particle Swarm Optimization algorithms to classical SVM algorithms [18–23]. The Particle Swarm Optimization (PSO) algorithm was developed by Eberhart and Kennedy and is a population-based stochastic optimization technique inspired by social behavior of bird flocking or fish schooling [24]. The PSO initializes the system with a population of random solutions and searches for optima by updating generations. The particles are the potential solutions, by optimizing the particles to find the best particle is the optimal solution to the problem. Compared with other stochastic optimization techniques like Genetic Algorithms (GAs), the advantages of PSO are that it is easy to implement and there are few parameters to adjust [25]. This technique can find the optimum parameters for the SVM model to better perform precipitation prediction. The PSO technique has been used widely in the optimization of SVM. For instance, Huang and Dun [18] proposed a

PSO-SVM model to improve classification accuracies with an appropriate feature subset. Zhao and Yin [20] take the global optimal search through PSO to obtain the SVM to effectively identify the geomechanical parameters. Selakova et al. [26] proposed a practical new hybrid model for short term electrical load forecasting based on PSO-SVM.

The PSO algorithm looks for optimal parameters in the prediction region for the SVM model. In this way, it can be more effective to overcome the problem of local optimum. Thus, the PSO-SVM model is used for precipitation prediction. A hybrid PSO optimized SVM (PSO-SVM) model was used as an automated learning tool, trained in order to predict other parameters such as optical density, dissolved oxygen concentration, pH, nitrate concentration and phosphate concentration [27]. The solving of SVM involves quadratic programming, and the computational complexity rapidly increases as the training samples increase in number. In this paper, SVM-PSO has outstanding classification precision. Therefore, this paper will focus on the SVM-PSO model and how to apply the model to precipitation prediction.

The remainder of this paper is organized as follows: in the following section, an introduction to SVM and PSO is presented. Section 3 presents the proposed SVM-PSO algorithm. In Section 4, we describe the features of data used for prediction of precipitation. Section 5 reports experiments including comparative results between the traditional SVM-based prediction models and ours. Section 6 presents some discussions of our model and the last section states our conclusions.

2. Algorithm

2.1. SVM

With the development of society, people are increasingly demanding higher accuracy of weather forecasts. Improving the accuracy of forecasting is an inevitable trend in forecasting business development. In weather forecast work, the knowledge accumulation and approximate memory of the work forecast model based on different types of weather are still used [28]. The accumulation of forecasting experience is a long-term process, and the complexity and non-linearity of weather evolution makes it difficult to identify the knowledge of forecasting. The evolution of any weather or meteorological elements is essentially the result of a combination of conditional combinations of certain meteorological elements, and the combination of these factors is varied and complex. Thus, in order to express complex nonlinear relationships between meteorological elements in specific time and space with advances in computer technology and intelligent technology, intelligent machine identification skills have been well developed. Support Vector Machine (SVM) is widely used for many machine learning problems. For example, the literature [29,30] used M-SVM and GA-SVM. SVM has many unique advantages in solving small sample, nonlinear and high dimensional pattern recognition problems. It is given by Vapnik in 1995 [31]. SVM is a classification method based on statistical learning and Vapnik–Chervonenkis (VC) dimensional theories [32]. The basic principle is briefly introduced as follows in this section.

Given separable sample sets $(\mathbf{x}_i, y_i)_{1 \leq i \leq N}$ as training samples, N is the total number of vectors. Training vectors $\mathbf{x}_i \in R^D$ are the input vectors and D is the dimension of the input space. y_i denotes the corresponding output of $\mathbf{x}_i$, $y_i = \pm 1$, which is denoted as precipitation occurrence or not here. Separation of the hyperplane is $w \cdot \mathbf{x} + b = 0$, where $(\cdot)$ is the inner product. w is the weight vector and b denotes the bias term. Equation (1) represents the classification hyperplane. Any of the training tuples that fall on the hyperplane form the equal sign of Equation (1), called the support vector:

$$y_i(w \cdot \mathbf{x}_i + b) \geq 1 \tag{1}$$

In order to maximize the interval, one only needs to maximize $\|w\|^{-1}$, which is equivalent to minimizing $\frac{1}{2}\|w\|^2$. The basic type of SVM is as shown in Equation (2):

$$\min_{w,b} \frac{1}{2}\|w\|^2$$
$$subject \, to \, y_i(w \cdot \mathbf{x}_i + b) - 1 \geq 0, \tag{2}$$

where $i = 1, 2, \cdots, N$. According to Equation (2), we can obtain a hyper-plane $w \cdot \mathbf{x}_i + b = 0$ with the largest margin. For equality constraints in Equation (2), they can be unconstrained by using Lagrangian, and we introduce the soft margin, in order to make a soft margin adapt to noisy data if the case is undivided, leading to:

$$\min_{w,b} \frac{1}{2}\|w\|^2 - C \sum_{i=1}^{N} \ell_{0/1}\left[y_i(w \cdot \mathbf{x}_i + b) - 1\right]. \tag{3}$$

$\ell_{0/1}$ is the "0/1 loss function". C is a constant greater than 0, and, when C is infinite, we require all samples to meet the constraints; otherwise, we allow some samples to not meet the constraints.

In this case, the dual problem of the objective function is a maximized quadratic programming problem. It can be defined as Equation (4):

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j), \tag{4}$$

Subject to:

$$\sum_{i=1}^{N} \alpha_i y_i = 0,$$
$$0 \leq \alpha_i \leq C, i = 1, 2, \cdots, N. \tag{5}$$

This is an extremum problem of quadratic function with inequality constraints, and there exists a unique solution. a_i is a Lagrange multiplier for each training sample, the sample is the support vector for which $a_i \neq 0$, lying on one of the two hyper-planes: $(w \cdot \mathbf{x}^+) + b = +1$; $(w \cdot \mathbf{x}^-) + b = -1$. According to the Karush–Kuhn–Tucker (KKT) condition, the optimization problem must satisfy the last one in Equation (5) [33]. When dealing with the nonlinear SVM problem, SVM introduces a kernel function to map the data and make data into a high dimensional space:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j). \tag{6}$$

The $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function [34]. $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$. When the training algorithm appears $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$, we can use the $K(\mathbf{x}_i, \mathbf{x}_j)$ instead. The constraint is the same as Equation (4), and $0 \leq \alpha_i \leq C$. SVM is based on the study of the kernel. SVM is a kernel-based learning algorithm, the key to the algorithm is the kernel function, so the kernel function has an important position, and the selection of kernel function affects the generalization ability of the model directly. Commonly used kernel functions are polynomial kernel functions, the RBF function, etc. The Gaussian Radial Basis Function (RBF) is one of the most popular kernel functions, and the RBF kernel function is accounted for in the nonlinear problems in this paper:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-g\|\mathbf{x}_j - \mathbf{x}_i\|^2\}, \tag{7}$$

where g is the kernel parameter to measure the width of kernel function in RBF. If g is inappropriate, the results of the algorithm may outfit or overfit the training data. This function is robust and can account for the nonlinear decision boundary.

Based on the above analysis, the basic idea of SVM is that, when classifying the nonlinear samples, the samples of the original space are mapped to a high-dimensional feature space by a nonlinear mapping, so that the space becomes linear and then linearly classifies in the high-dimensional space, so as to realize the relative original space nonlinear algorithm.

2.2. Particle Swarm Optimization

As a popular parameter optimization tool, Particle Swarm Optimization (PSO) is also applied in this paper. Besides SVM, PSO is introduced to obtain the optimal parameters in SVM. It is an evolutionary computation technique [24] that was inspired by the social activity of birds primarily and then established a simplified model based on birds searching for food. Birds in a swarm cooperate with each other like swarm intelligence to search for the optimal position in order to obtain the food. The group of birds whose amount is m, called a "particle", form a population in a D-dimensional target search space. If the solution space of the optimization problem is D dimensions, the space vector $X_i = (x_{i1}, x_{i2}, x_{i3}, \cdots x_{iD})$ is represented as the *i*-th particle, where $i = 1, 2, \cdots, m$, X_i is the position of the *i*-th particle and also a possible solution.

The velocity and position of the particles are iterated to obtain the equation as follows [35]:

$$v_{id}^{t+1} = \omega \cdot v_{id}^t + c_1 \cdot r_1 \cdot (p_{id} - x_{id}^t) + c_2 \cdot r_2 \cdot (p_{gd} - x_{id}^t), \tag{8}$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}, \tag{9}$$

where $V_i = (v_{i1}, v_{i2}, v_{i3}, \cdots v_{iD})$ is the velocity of the *i*-th particle, and $P_i = (p_{i1}, p_{i2}, p_{i3}, \cdots p_{iD})$ is the optimal position of this particle. The optimal swarm position is $P_g = (p_{g1}, p_{g2}, p_{g3}, \cdots p_{gD})$. Under the condition of the *i*-th particle at the *t*-th iteration, x_{id}^t and v_{id}^t are the *d*-th location and speed component. Positive coefficient c_1, c_2, r_1 and r_2 are the random number, the range is 0 to 1, and ω is the inertial weight of PSO algorithm.

The PSO algorithm can deal with the continuous optimization problem and support multi-point search. Therefore, we take the PSO algorithm determine the parameters of SVM to improve the performance of SVM model.

3. SVM Based on the PSO

Considering the influence of training parameters on generalization performance of SVM, we introduce the particle swarm optimization algorithm to search the training parameters in global space. In recent years, people have developed a series of methods based on the nuclear method of learning, and the linear learners expand into a nonlinear learning by using kernel functions. However, the problem of selecting parameters of a kernel function still exists [18]. The selection of parameters C and g have important influence on the operation results of the SVM model. Different kernel functions will produce different classification results for the same classification problem. Due to the lack of theoretical guidance, the choice of the traditional kernel function mostly depends on the repeated experiments and selects the satisfactory solution manually.

The parameter space exhaustive search method (enumeration method) is the most widely used SVM parameter optimization method, with each parameter in a certain interval in a certain range of values, each value in a level, and each parameter in a different level combination constitutes a combination of multiple sets of candidate parameters, and then selects the minimum expected risk from the set of a minimum level as the optimal parameter value. At the same time, the method combines *k*-fold cross validation and the enumeration method to select the parameters of SVM because it is the most accurate estimate on the upper bound risk of the SVM.

However, this method also has shortcomings:

(1) Time-consuming—when the data size becomes large, or the number of parameters exceeds two, it takes a very long time to calculate.
(2) It is difficult to find the optimal parameters accurately. The method needs to select a reasonable range for parameters, and the optimal parameters may not be in the range. For the practical problem, the above method can only find the local optimal solution in the candidate parameter combination.

Combination of the PSO algorithm and SVM model can effectively solve this problem. We use the PSO algorithm to optimize the parameters in this paper. Because the algorithm is easy to understand, the realization is simple, the adjustable parameters are few, and it is better than common intelligent algorithms like artificial neural network (ANN) and Genetic algorithm (GA) in many cases. Based on the previous discussion, taking the Gaussian Radial Basis Function (RBF) function as the kernel function, we demonstrate the flowchart of the PSO-SVM algorithm in Figure 1. As Figure 1 shows, The PSO-SVM process can be stated as the following steps: (1) Data to preliminary processing. This optimization method needs a training process based on the known data set to obtain the SVM model. Thus, we should generate the training set and test set randomly, while the data should be normalized and dimensioned; (2) initialize parameters of particle swarm optimization like population size, the maximum iteration number, and so on. Of course, also its parameters C and g are included; (3) particle swarm optimization is adopted to search the optimal solution of particles in global space by using a cross-validation algorithm; (4) put the training data into the SVM model with the optimal parameters to obtain the trained PSO-SVM model; and (5) the testing data is used to test the prediction ability of the trained PSO-SVM model.
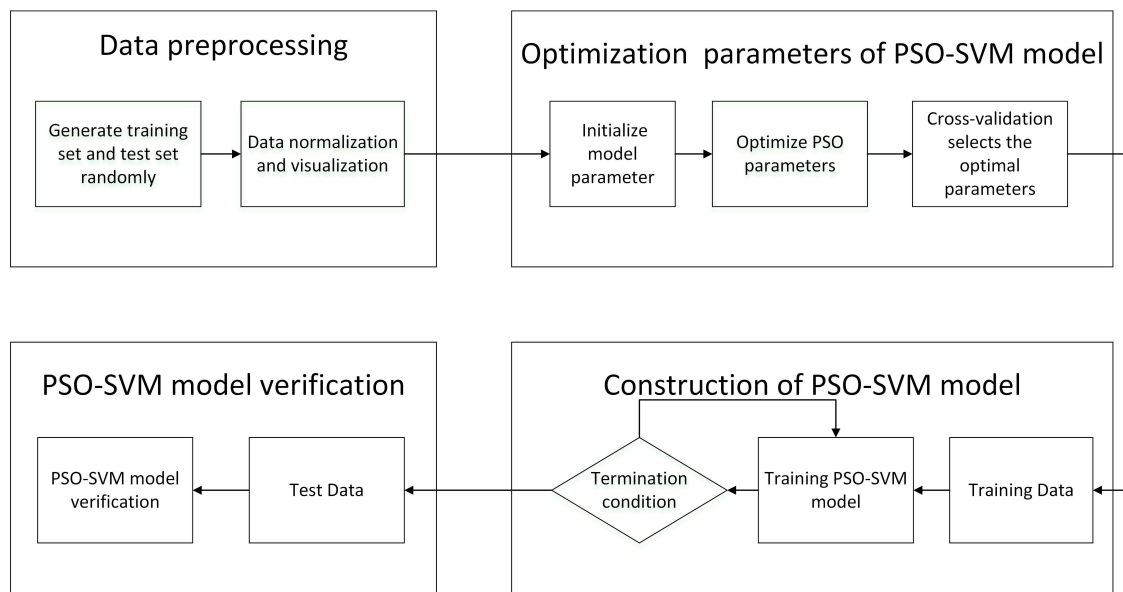

Figure 1. The flowchart of the Support Vector Machine with Particle Swarm Optimization model.

The details of PSO for selecting parameters for the SVM model can be stated as the following steps:

Step 1: Initialization parameters

Initialize parameters of particle swarm optimization like population size, the maximum iteration number and so on. The initial particle swarm location is x_i^0 and velocity v_i^0 should first be determined. We set $X = \text{rand}(-x_{id\,\max}, x_{id\,\max})$, where rand () is a random function in the range [0, 1].

Step 2: Evolution starts

Set gen = 1, produce randomly a set of *C* and *r* in the given valuing ranges. Each selected kernel function and its parameters such as *C* and *r* are regarded as an individual of SVM. Give the random position and the flight velocity for food finding of an individual particle *i*.

Step 3: Preliminary calculations

Calculate the new position x_i, v_i^0 of the *i*-th particle, and then input x_i into the SVM model for precipitation forecasting; the parameters $[C, g]$ of the SVM model are replaced by $[X_i(1), X_i(2)]$. According to the precipitation forecasting result, the f^i can be calculated. The f^i is represented by the *k*-fold cross-validation, which calculates the errors between the actual values and the forecasted values.

Step 4: Offspring generation

The global best value is generated according to Equations (8) and (9) after updating the position value. Then, the global best value is inputted into the SVM model and the fitness function value is calculated again. Set gen = gen + 1.

Step 5: Circulation stops

When gen is equal to the maximum iteration number, and the stop criterion is met, the optimal parameters of the SVM model are finally obtained. Otherwise, it is necessary to go back to Step 2. Then, end the training and verification procedure and get the result of the PSO-SVM model. The diagram of the SVM-PSO forecasting model is illustrated in Figure 2.
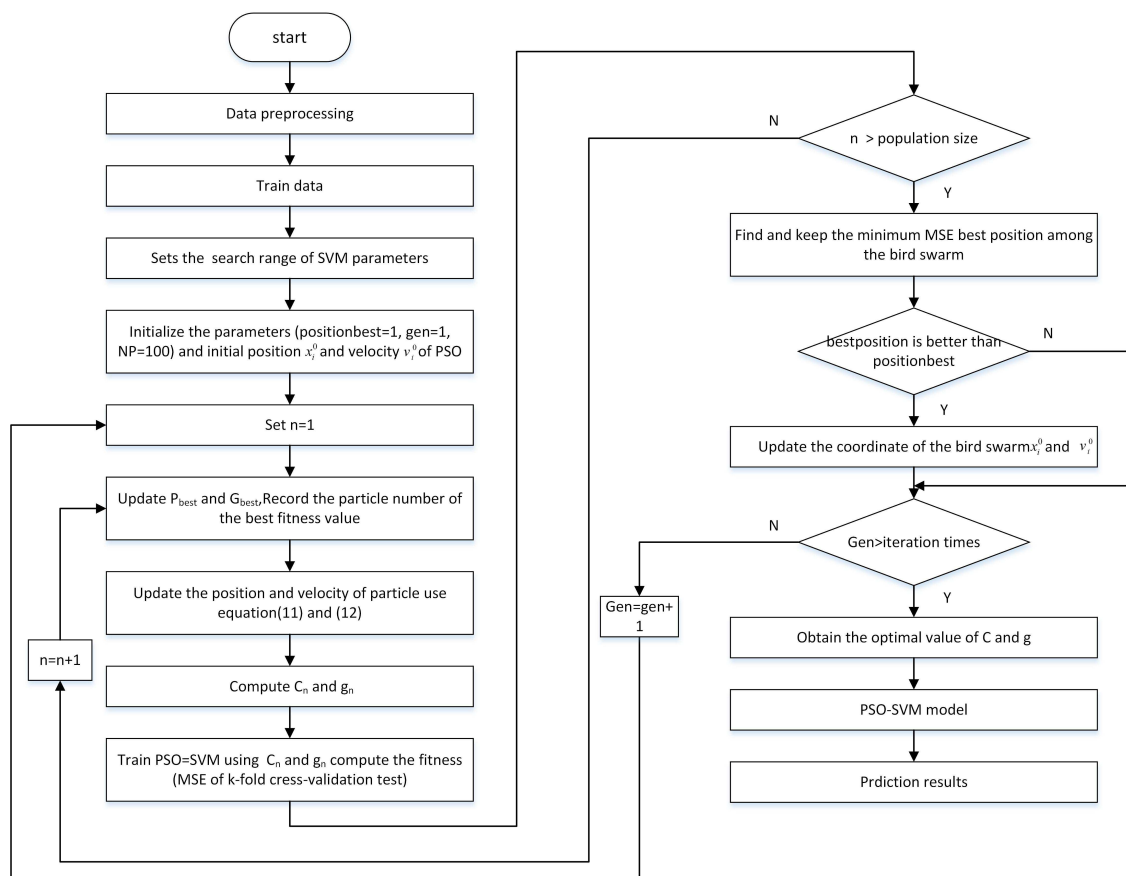

Figure 2. Diagram of the Support Vector Machine and Particle Swarm Optimization (SVM-PSO) forecasting model.

4. Experiments

4.1. Data Collection and Preprocessing

Data collection and preprocessing are the initial stages of the data mining process. Because only valid data will produce accurate output, data preprocessing is the key stage. For this study, we use the one-year ground-based meteorological data from Nanjing Station (ID: [58238]). The dataset contains atmospheric pressure, sea level pressure, wind direction, wind speed, relative humidity and precipitation. Data is collected every 3 h. We just consider related information and ignore the rest. The method of principal component analysis (PCA) is used to reduce the dimensionality of the data, thus reducing the data processing time and improving the efficiency of the algorithm. We performed data transformation on rainfall.

By observing the original data set, we find the incorrect data in the data set, which does not correspond to the fact. For example, if the pressure value of a sample is 999,017 or 999,999 and it appears repeatedly, it indicates that such erroneous data is abnormal data returned when data is acquired, so we reject samples containing such data. The error data is the data returned when the abnormality is collected. Therefore, we exclude the samples containing such data and check other elements until there is no data value that exceeds the normal threshold.

As shown in Table 1, the first line is the attributes of the original meteorological data set; PRS represents atmospheric pressure, PRS_Sea is the sea level pressure, WIN_D is the wind direction, WIN_S is the wind speed, TEM is the temperature, RHU is the relative humidity, and PRE _1h is the hourly precipitation.

Table 1. Original meteorological data.

PRS (hPa)	PRS_Sea (hPa)	WIN_D (°)	WIN_S (0.1 m/s)	TEM (°C)	RHU (1%)	PRE_1h (mm)
1031.2	1035.8	89	2.5	77	2	0
1030.8	1035.4	113	2.9	61	6.4	0
1027.3	1031.9	153	2.1	49	8.3	0
1026.2	1030.8	122	2	55	7.1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1027.1	1031.7	121	0.7	71	4.1	0

The statistical performance of the classification model can be evaluated by comparing the labeled data sets. The label attribute stores the actual observed values, and the prediction attribute stores the values of labels predicted by the classification model. For the assessment of the algorithms' results, calibration, cross-validation and external testing were carried out. For each one of the experiments, its datasets were divided into two subsets, training and test, comprising 80% and 20% of the original samples, respectively. Thus, we randomly classify datasets into training sets and test sets.

I.   Training set—take 80% of the samples randomly from the dataset as the training set.
II. Text set—for the text data set, we used other data remaining in the data set, which contained all the attributes except the rainfall data that the model is supposed to predict. The test set was never used for the training of any of the models.

In both two sets, the attributes and their types remained the same.

While the data are classified according to different amounts of rainfall, the data is divided into 0 categories when the data of the rainfall is 0, and the rainfall of data is non-0 divided into one category. There is a total of 2214 pieces of valid data, of which 1974 of data belong to 0 categories, accounting for 90% of the total data, 240 pieces of data belong to one category, accounting for 10% of the total data.

4.2. Data Normalization

In general, the data need to be normalized when the sample data are scattered and the sample span is large, so that the data span is reduced to build the model and prediction. In the SVM modeling, in order to improve the accuracy of prediction and smooth the training procedure, all of the sample data are normalized to fit them in the interval [0,1] using the following linear mapping formula $(-x_{id\,\max}, x_{id\,\max})$:

$$X' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad i = 1, 2, 3 \cdots, N, \tag{10}$$

where X' is the mapped value; x is the initial value from the experimental data; N is the total number; x_i is i of input data; $x_{\max}$ and $x_{\min}$ denote the maximum and minimum values of initial data, respectively.

4.3. Algorithm Validation

In order to verify the predictive ability of the precipitation prediction model based on the proposed PSO-SVM algorithm, the model is compared with the traditional SVM, GA-SVM and the AC-SVM (SVM with Ant Colony Algorithm) precipitation prediction model. We call it the GA-SVM method, when the SVM algorithm uses the Genetic Algorithm to select and optimize the parameters. Similarly, the AC-SVM model optimizes the parameters of SVM by Ant Colony Algorithm, and the M-SVM model is the SVM model which uses the numeration method to select the parameters. Many scholars used the GA-SVM model and AC-SVM in forecasting problems and achieved good predictive performance [36–40]. The M-SVM model using the numeration method is a widely used traditional SVM method. These four models are the same in both the training sample and the predicted sample in the comparison test.

For the assessment of the algorithms' results, calibration, cross-validation and external testing were carried out. In the calibration assessment, the models were developed using the training set and validated with the same one. In the cross-validation, a k-fold method was performed upon the training dataset with a k value of 5 (in order to maintain the 80:20 ratio, obtaining five executions where, in each one of them, the model is trained with 80% of the samples and evaluated testing the remaining 20%). Finally, the prediction results were obtained via an external validation, training and testing the models with the training and test datasets, respectively.

In this paper, the number of samples is 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, for each type of sample number, and 80% of the data were randomly selected as the training set for constructing the algorithm model. The remaining 20% of the data were used as the test set to validate the model accuracy. According to the trial, the sensitivity coefficient was $\xi = 0.001$, using three different optimization algorithms to optimize the penalty factor C and the kernel function width coefficient g.

The coefficient of variation (CV) is defined as the ratio of the standard deviation. When you need to compare two sets of data when the size of the degree of dispersion and measurement scales are much different for the two sets of data, or the data for different dimensions, direct use of standard deviation for comparison is inappropriate, and you must eliminate measurement scales and dimensions. In addition, the coefficient of variation can do this, which is the ratio of the standard deviation of the original data to the mean of the original data.

The smaller the coefficient of variation, the smaller the variation (deviation) level—on the contrary, the greater the risk. The abscissa is the number of samples, and the ordinate is the accuracy. The accuracy of the traditional mesh optimization method is more stable, but it is easy to fall into the local optimal value. In the case of small sample size, the accuracy of GA parameter optimization is higher, but the stability is bad, which is influenced by the sample. The accuracy of Ant Colony optimization is the most stable among the four optimization methods, which is about 92%. Changes in the sample size have minimal impact on accuracy. PSO parameter optimization method has the best accuracy in the case of small sample size. With the increase of the number of samples, the accuracy rate increases and tends to be stable. The SVM algorithm model trained by the training set contains the best RBF kernel parameter g and the penalty parameter C, removes the label in the training set,

and then puts it into the model for prediction, and the comparison between the prediction result and the actual tag is carried out, and the accuracy curve of the model with the size of the training set under different optimization parameters is obtained. It can be seen from the Figure 3 that, since the model is based on this data set, the accuracy rate is quite high, and the accuracy of some data is close to 100%. However, with the increase of data size, the accuracy rate has declined and has become unstable.
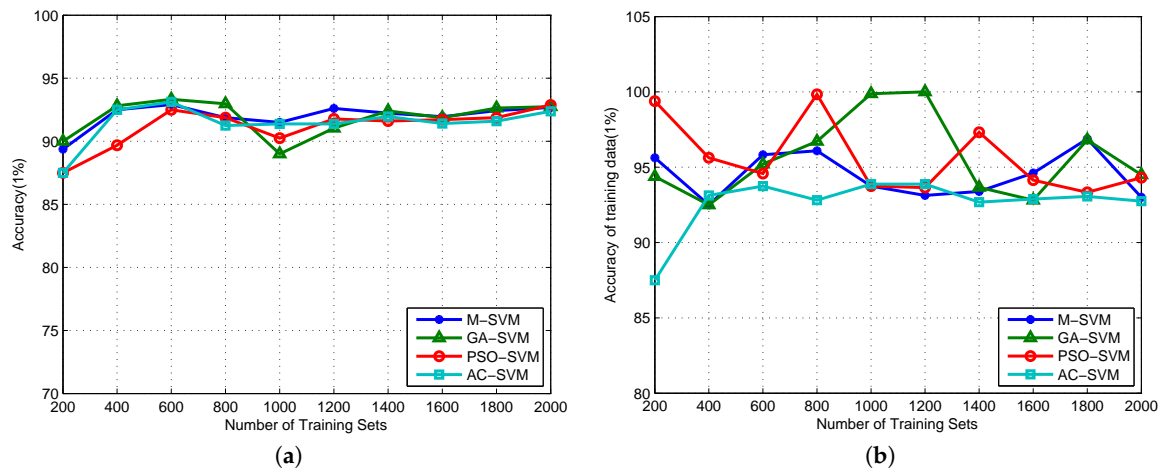

Figure 3. (**a**) the best classification accuracy curve under the final coefficient of variation (CV) condition is the training set based on different parameters' optimization algorithms; (**b**) the curve of accuracy rate changes with the data scale by using a model of training data to predict the training set.

The prediction accuracy fluctuates greatly when the number of samples is small. The accuracy rate reaches the lowest point when the number of samples is 600, with the increase in the number of samples, the accuracy rate increases, the accuracy based on M-SVM reaches the maximum when the number of samples is 1200, and then the accuracy rate begins to decline. The GA-SVM model reaches the maximum when the number of samples is 1400, and then the accuracy rate begins to decline. The AC-SVM model gets the best classification accuracy when the number of samples is 800, but then it has poor performance. The accuracy based on the PSO optimization method achieves the maximum value when the number of samples is 1600, and the accuracy is basically better than the GA method. We just used the SVM method and the training set established a model without optimization of the parameter selection. The accuracy of the test data based on this SVM model is shown in Figure 4. With reference to SVM, the GA-SVM and PSO-SVM models are more accurate when the number of samples is large.
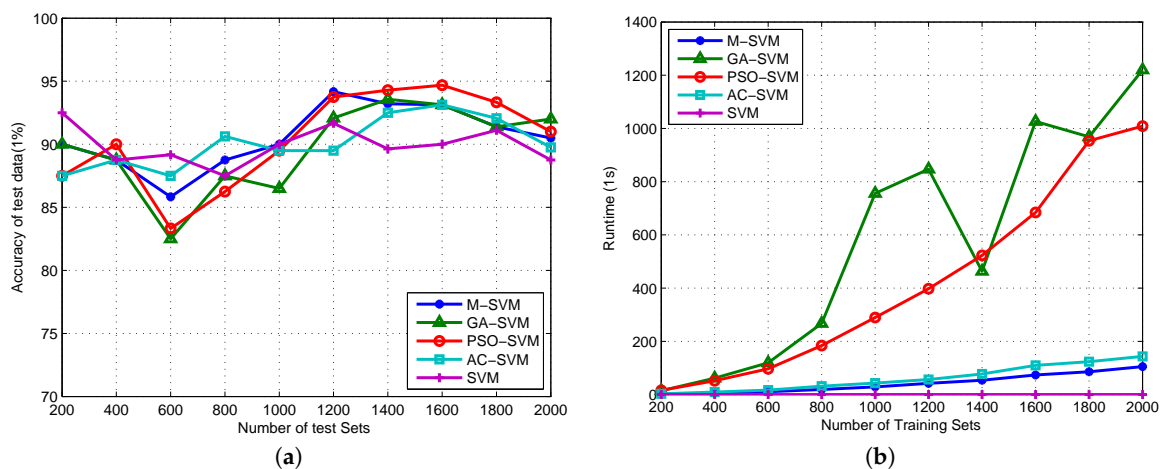

Figure 4. (**a**) the accuracy curve obtained by running the test data in the trained model; (**b**) the time required to build the model as the number of sample changes.

The model without optimization parameters takes less time to establish and can be neglected. The time curves of M-SVM and AC-SVM are linear, the slopes are small and the growth is slow. As the number of samples increases, the curve of the PSO-SVM method changes faster, and the time is positively correlated with the sample size. The curve of GA-SVM is on the rise, and the pre-growth rate is the fastest among the four methods. With the increase in the number of samples, the time required to establish the model becomes extremely unstable. Overall, the GA-SVM method takes the longest time to calculate.

5. Discussion

First of all, for data preprocessing, we need to remove the bad datum and abnormal values before attempting to calculate, then normalize the data to eliminate the effects of sample spans, and smooth the training process. The data is subjected to a fivefold cross validation for a more stable data model. Due to the particularity of the precipitation datum, the regional precipitation datum are distributed unevenly in time and space. The days of precipitation are obviously shorter than that of the total sample. Although the datum are pretreated, they still cannot completely remove the datum on the impact of the forecast results, and the degree of volatility is greater when the number of samples is small.

We analyze the results of the optimization parameters C and g with the four methods. The PSO-SVM method based on the Particle Swarm Optimization algorithm has better experimental results, with a higher classification rate among the four parameters' optimization methods, which are the traditional M-SVM and the GA-SVM based on the Genetic algorithm, and the AC-SVM based on the Ant Colony algorithm. This indicates that the Particle Swarm Optimization algorithm has a strong ability to optimize the parameters of the Support Vector Machine C and g, and the results are more accurate.

In order to further analyze the performance problem of PSO-SVM based on the Particle Swarm Optimization algorithm and the algorithms based on bionic intelligence GA-SVM and AC-SVM, we use the data set whose sample size is 1000 as an example, and the generalization ability of these four algorithms is shown in Figure 5. In Figure 5, the abscissa is the FPRate (False positive rate) and the ordinate is the TPRate (true positive rate). The curve in Figure 5 is the ROC (Receiver Operating Characteristic) curve drawn from the wired sample. The two learners are judged by comparing the area under the curve, that is, the AUC (Area Under Curve). Secondly, we use MSE (Mean Square Error) to evaluate the stability of the four methods. It can be seen from the experimental results in Table 2 that the four methods are different for different samples of MSE, but on the whole, the SVM model parameters are selected by using the Particle Swarm Optimization algorithm, which is much more stable than the Genetic algorithm and the Ant Colony algorithm. With small samples, the accuracy of the PSO-SVM method and the GA-SVM method are considerable, but the PSO-SVM is more efficient, and, with the increase in the number of samples, PSO-SVM will be more stable and efficient. Compared with AC-SVM, PSO-SVM has higher classification accuracy; however, the computation time of the PSO-SVM model is slightly longer with the large data set. These four algorithms are better for large data classification. It can be seen from Table 2 and Figures 4 and 5 that the SVM based on Particle Swarm Optimization algorithm has better learning performance. Experiments show that SVM model parameter selection based on the Particle Swarm Optimization is effective and feasible.
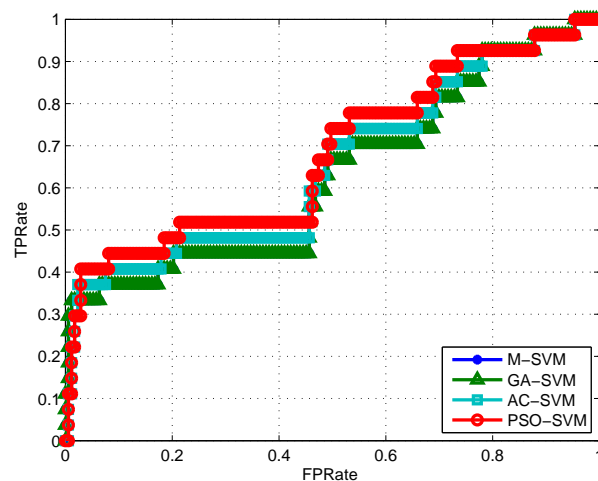
Figure 5. The Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC) drawn from the wired 1000 samples.

Table 2. The Mean Square Error (MSE) of four methods.

MSE	M-SVM	GA-SVM	AC-SVM	PSO-SVM
200	0.3162	0.3162	0.3536	0.3536
400	0.4183	0.3354	0.3354	0.3354
600	0.3873	0.3416	0.3536	0.3651
800	0.3708	0.3708	0.3062	0.3354
1000	0.3162	0.324	0.324	0.3162
1200	0.2814	0.25	0.25	0.2415
1400	0.2892	0.2809	0.2739	0.2597
1600	0.2622	0.25	0.2622	0.2305
1800	0.2934	0.2635	0.2635	0.2582
2000	0.3082	0.2872	0.2872	0.2782

6. Conclusions

The PSO-SVM algorithm is proven to be an effective method of the rainfall forecast decision. The SVM method is a kind of machine learning method with a high degree of nonlinear problems. It is a kind of machine intelligent learning method with a solid theoretical basis. Moreover, there is no limit to the dimension (number of vectors) of the vector due to the establishment of the SVM model, which facilitates the handling of meteorological problems with time, space and multiple factors. We modeled the parameters of the model by using the PSO algorithm to optimize the parameters of the model. The PSO-SVM model was established and compared with the traditional mesh optimization, the Genetic algorithm and the Ant Colony algorithm.

The experimental results indicate that the PSO algorithm has a higher accuracy and efficiency. The PSO-SVM is ideal for multiple variable analyses and particularly important in current problem-solving tasks like weather forecasts. Furthermore, some research could be done to increase the accuracy rate of this model. The output of the model is proven to be very promising and absolutely helpful for its improvement.

Thus, this study is considered significant. If the weather warnings are told to ordinary people ahead of the natural calamity, they can act as life-saving signals, valuable added services and disaster prevention tools. These models provide a good example of the capabilities of Support Vector Machine Regression for modeling high precision and efficient weather forecasting.

Author Contributions: J.D. and Y.L. conceived and designed the experiments; Y.Y. and W.Y. performed the experiments; J.D. and Y.L. analyzed the data; Y.L. contributed analysis tools; J.D. and Y.L. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Geetha, A.; Nasira, G.M. Data mining for meteorological applications: Decision trees for modeling rainfall prediction. In Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 18–20 December 2014; pp. 1–4.

2. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Elsevier: Harrisburg, PA, USA, 2011; ISBN 10: 9380931913.

3. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Witten, I.H., Frank, E., Eds.; Elsevier: Burlington, MA, USA, 2014; ISBN 9787111453819.

4. Yin, Y.; Zhao, Y.; Li, C.; Zhang, B. Improving Multi-Instance Multi-Label Learning by Extreme Learning Machine. *Appl. Sci.* **2016**, *6*, 160, doi:10.3390/app6060160.

5. Annas, S.; Kanai, T.; Koyama, S. Assessing daily tropical rainfall variations using a neuro-fuzzy classification model. *Ecol. Inform.* **2007**, *2*, 159–166, doi:10.1016/j.ecoinf.2007.04.001.

6. Prasad, N.; Reddy, P.K.; Naidu, M.M. A Novel Decision Tree Approach for the Prediction of Precipitation Using Entropy in SLIQ. In Proceedings of the 2013 UKSim 15th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, 10–12 April 2013; pp. 209–217.

7. Lu, K.; Wang, L. A novel nonlinear combination model based on support vector machine for rainfall prediction. In Proceedings of the IEEE 4th International Joint Conference on Computational Sciences and Optimization (CSO 2011), Kunming and Lijiang City, China, 15–19 April 2011; pp. 1343–1347.

8. Kisi, O.; Cimen, M. Precipitation forecasting by using wavelet-support vector machine conjunction model. *Eng. Appl. Artif. Intell.* **2012**, *25*, 783–792, doi:10.1016/j.engappai.2011.11.003.

9. Ortiz-Garcia, E.G.; Salcedo-Sanz, S.; Casanova-Mateom, C. Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. *Atmos. Res.* **2014**, *139*, 128–136, doi:10.1016/j.atmosres.2014.01.012.

10. Sanchez-Monedero, J.; Salcedo-Sanz, S.; Gutierrez, P.A.; Casanova-Mateo, C.; Hervas-Martinez, C. Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-rdinal support vector classifier. *Eng. Appl. Artif. Intell.* **2014**, *34*, 199–207, doi:10.1016/j.engappai.2014.05.016.

11. Sehad, M.; Lazri, M.; Ameur, S. Novel SVM-based technique to improve rainfall estimation over the Mediterranean region (north of Algeria) using the multispectral MSG SEVIRI imagery. *Adv. Space Res.* **2017**, *59*, 1381–1394, doi:10.1016/j.asr.2016.11.042.

12. Young, C.C.; Liu, W.C.; Wu, M.C. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Appl. Soft Comput.* **2017**, *53*, 205–216, doi:10.1016/j.asoc.2016.12.052.

13. Wei, J.; Huang, Z.; Su, S.; Zuo, Z. Using Multidimensional ADTPE and SVM for Optical Modulation Real-Time Recognition. *Entropy* **2016**, *18*, 30, doi:10.3390/e18010030.

14. Li, W.; Huang, Z.; Lang, R.; Qin, H.; Zhou, K.; Cao, Y. A Real-Time Interference Monitoring Technique for GNSS Based on a Twin Support Vector Machine Method. *Sensors* **2016**, *16*, 329, doi:10.3390/s16030329.

15. Sonnenschein, A.; Fishman, P.M. Radiometric detection of spread-spectrum signals in noise of uncertain power. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 654–660, doi:10.1109/7.256287.

16. De Giorgi, M.G.; Campilongo, S.; Ficarella, A.; Congedo, P.M. Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN). *Energies* **2014**, *7*, 5251–5272, doi:10.3390/en7085251.

17. Xiao, C.C.; Hao, K.R.; Ding, Y.S. The bi-directional prediction of carbon fiber production using a combination of improved particle swarm optimization and support vector machine. *Materials* **2015**, *8*, 117–136, doi:10.3390/ma8010117.

18. Huang, C.L.; Dun, J.F. A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Appl. Soft Comput.* **2008**, *8*, 1381–1391, doi:10.1016/j.asoc.2007.10.007.

19. Ren, F.; Wu, X.; Zhang, K.; Niu, R. Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. *Environ. Earth Sci.* **2015**, *73*, 4791–4804, doi:10.1007/s12665-014-3764-x.

20. Zhao, H.; Yin, S. Geomechanical parameters identification by particle swarm optimization and support vector machine. *Appl. Math. Model* **2009**, *33*, 3997–4012, doi:10.1016/j.apm.2009.01.011.

21. Zhang, Y.; Wu, L. Classification of fruits using computer vision and a multiclass support vector machine. *Sensors* **2012**, *12*, 12489–12505, doi:10.3390/s120912489.

22. Zhang, Y.D.; Dong, Z.C.; Liu, A.J.; Wang, S.H.; Ji, G.L.; Zhang, Z.; Yang, J.Q. Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. *J. Med. Imaging Health Inform.* **2015**, *5*, 1395–1403, doi:10.1166/jmihi.2015.1542.

23. Wu, J.; Liu, M.; Jin, L. A hybrid support vector regression approach for rainfall forecasting using particle swarm optimization and projection pursuit technology. *Int. J. Comput. Intell. Appl.* **2010**, *9*, 87–104, doi:10.1142/S1469026810002793.

24. Kennedy, J.; Eberhart, R. Particle swarm optimization. *IEEE Proc. Int. Conf. Neural Netw.* **1995**, *4*, 1942–1948; ISBN 0-7803-2768-3.

25. Hassan, R.; Cohanim, B.; De Weck, O.; Venter, G. A comparison of particle swarm optimization and the genetic algorithm. In Proceedings of the 46th AIAA Multidisciplinary Design Optimization Specialist Conference, Austin, TX, USA, 18–21 April 2005; pp. 18–21.

26. Selakov, A.; Cvijetinovic, D.; Milovic, L.; Mellon, S.; Bekut, D. Hybrid PSO-SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank. *Appl. Soft Comput.* **2014**, *16*, 80–88, doi:10.1016/j.asoc.2013.12.001.

27. Shieh, M.Y.; Chiou, J.S.; Hu, Y.C.; Wang, K.Y. Applications of PCA and SVM-PSO based real-time face recognition system. *Math. Probl. Eng.* **2014**, *2014*, doi:10.1155/2014/530251.

28. Nong, J.F.; Jin, L. Application of support vector machine to predict precipitation. In Proceedings of the 7th World Congress on Intelligent Control and Automation, Chingqing, China, 25–27 June 2008; pp. 8975–8980.

29. Wang, S.H.; Yang, X.J.; Zhang, Y.D.; Phillips, P.; Yang, J.F.; Yuan, T.F. Identification of green, oolong and black teas in China via wavelet packet entropy and fuzzy support vector machine. *Entropy* **2015**, *17*, 6663–6682, doi:10.3390/e17106663.

30. Zhang, Y.D.; Dong, Z.C.; Wang, S.H.; Ji, G.L.; Yang, J.Q. Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *Entropy* **2015**, *17*, 1795–1813, doi:10.3390/e17041795.

31. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 1st ed.; Springer-Verlag: New York, NY, USA, 1995; ISBN 978-1-4757-2442-4.

32. Vapnik, V.N.; Chervonenkis, A. *On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities*; Theory of Probability and Its Applications; Springer International Publishing: New York, NY, USA, 1971; pp. 264–280; ISBN 978-3-319-21851-9.

33. Kuhn, H.; Tucker, A. Nonlinear Programming. In Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics, Berkeley, CA, USA, 31 July–12 August 1950; pp. 481–492.

34. Ballanti, L.; Blesius, L.; Hines, E.; Kruse, B. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sens.* **2016**, *8*, 445, doi:10.3390/rs8060445.

35. Kennedy, J. Particle Swarm Optimization. In *Encyclopedia of Machine Learning*; Springer: Berlin, Germany, 2011; pp. 760–766; ISBN 978-0-387-30768-8.

36. Chi, D.C.; Zhang, L.F.; Xue, L.I.; Wang, K.; Xiu-Ming, W.U.; Zhang, T.N. Drought Prediction Model Based on Genetic Algorithm Optimization Support Vector Machine (SVM). *J. Shenyang Agric. Univ.* **2013**, *2*, 013, doi:10.3969/j.issn.1000-1700.2013.02.011.

37. Zang, S.Y.; Zhang, C.; Zhang, L.J.; Zhang, Y.H. Wetland Remote Sensing Classification Using Support Vector Machine Optimized with Genetic Algorithm: A Case Study in Honghe Nature National Reserve. *Sci. Geograph. Sin.* **2012**, *4*, 007, doi:10.13249/j.cnki.sgs.2012.04.006.

38. Gao, L.F.; Zhang, X.L.; Wang, F. Application of improved ant colony algorithm in SVM parameter optimization selection. *Comput. Eng. Appl.* **2015**, *51*, 139–144, doi:10.3778/j.issn.1002-8331.1307-0281.

39. Niu, D.; Wang, Y.; Wu, D.D. Power load forecasting using support vector machine and ant colony optimization. *Expert Syst. Appl.* **2010**, *37*, 2531–2539, doi:10.1016/j.eswa.2009.08.019.

40. Ni, L.P.; Ni, Z.W.; Li, F.G.; Pan, Y.G. SVM model selection based on ant colony algorithm. *Comput. Technol. Dev.* **2007**, *17*, 95–98. (In Chinese)