PV Prediction based on PSO-GS-SVM Hybrid Model

KONG Hongmei¹,a, SUI Huibin¹,b and ZHANG Peng¹

¹Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan; Shandong Province; 250061, China

¹¹15954116782@163.com, b,suihuibin@mail.sdu.edu.cn

Abstract. Photovoltaic power generation is affected by many factors, with volatility and intermittent characteristics. Large-scale photovoltaic access to the power grid poses great challenges to the safety and stability of power systems. Therefore, accurate prediction of photovoltaic power generation helps dispatchers adjust scheduling schedules in a timely manner, effectively reducing the adverse impact of photovoltaic power generation access on the power grid. This paper proposes a hybrid PV power prediction model based on PSO-GS-SVM. The particle swarm optimization (PSO) method is used to optimize the large step size of the support vector machine (SVM), and the parameter optimization range is obtained. GridSearch Method (GS) refined parameters optimization of PSO-SVM, and obtained PSO-GS-SVM hybrid model. The model is used to train and predict the normalized and dimensional sunny and non-clear working conditions data sets, and compared with BP neural network, SVM and PSO-SVM models. The results show that the PSO-GS-SVM hybrid model has better generalization ability and higher fitting effect.

1. Introduction
The huge amount of solar energy, abundant resources, and clean and pollution-free, these PV power generation has been rapidly developed, and the global installed capacity has increased year by year. PV power generation is affected greatly by environmental factors such as weather, and its power generation has characteristics of intermittent and volatility. Large-scale PV access poses great challenges to the power system and seriously affects its safe and stable operation. Accurate forecasting of PV power generation will help grid personnel to develop reasonable power generation plans and optimize power supply combinations, effectively solve PV grid-connected problems, furthermore, improve power station economic operation and market competitiveness [1].

At present, there are several ways of PV prediction methods. According to different modeling methods, they can be divided into physical methods and statistical methods. Although the physical method does not require a large amount of historical data, it requires detailed geographic information and component parameters of the PV plant, the modeling process is complicated, while the model has poor anti-interference ability [2]. The statistical method is based on the statistical law between the input and output of the predictive model. The artificial intelligence algorithm includes neural network and SVM, the modeling is simple and easy to implement. The BP neural network prediction model is established with the historical power generation sequence, daily type index and temperature data of the PV array[3]. However, the network structure lacks a unified and complete theoretical guidance, and it is easy to fall into the local minimum value. The actual prediction effect is not good enough. The cross-validation method is used to determine the parameters of SVM, and the model based on PV
output power and time parameters is established, but the calculation cost is large[4]. When the data set is large, the running speed is very slow. The combined model prediction method based on set empirical mode decomposition and SVM has good predictive ability under both mutated weather and non-mutant weather, but Theil coefficient is higher than that of BP neural network model[5].

This paper proposes a PSO-GS-SVM hybrid model to predict the PV power of the following day. Firstly, the data set is normalized and reduced in dimension, and the feature quantity is selected. PSO algorithm is used to optimize the large-step parameters of SVM, and the optimization range is delineated. Then GS is used to refine the parameters. The model is used to train and forecast the sunny and non-clear working conditions data sets. In order to verify the prediction accuracy of the model, BP neural network, SVM and PSO-SVM models were established as comparison examples.

2. Feature selection
After obtaining the sample data, it is necessary to clear the abnormal data, such as the fault and the limit state, to ensure the accuracy of the prediction model. The prediction model based on machine learning method needs to normalize the data to reduce the difference of numerical range of different physical quantities. It can be seen from Figure 1 that the weather irradiance varies greatly depending on the type of weather. Therefore, the data set is divided into two categories: clear air conditions and non-clear air conditions. Because there are too many PV influence factors, in the research the Principal Component Analysis (PCA) is adopted to reduce the input data, for the sum of the first, second, third and fourth principal components can describe more than 98% of the information, as shown in Figure 2, which greatly improves the model running speed.

3. Method

3.1. SVM
SVM is an effective machine learning method based on statistical learning theory and the principle of structural risk minimization. The SVM maps the input vector to a high-dimensional feature space using a nonlinear map, and then uses a linear function to implement regression prediction. The SVM regression is a given sample data set \( \{(x_i, y_i)\} \) \( i=1,2,...,n \), which solves the optimization problem of the equation (1) [6]:

\[
\min \left\{ \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^{n} (\xi_i + \xi_i^*) \right\}
\]

Where \( \omega \) is the weight vector and \( c \) is the penalty factor. When \( c \) is too large, there will be a phenomenon of learning; \( \xi \) is a slack variable. The Lagrange function is introduced to convert it into a dual problem, and its kernel function \( K(x, y) \) satisfies the Mercer condition. This research uses the radial basis kernel function of equation (2), where the kernel function parameter \( g \) affects the complexity of the model:

\[
K(x, y) = \exp \left( -\frac{\|x-y\|^2}{2g^2} \right)
\]
\[ K(x_i, y_i) = \exp\left(-g\|x_i - y_i\|^2\right), g > 0 \] (2)

SVM faces the problem of model selection in practical applications. SVM model includes the selection of kernel functions, and the penalty factor. This research chooses the \( c \) and \( g \) as the optimization objects. By continuously trying the combination of \( c \) and \( g \), the parameters with the best fitting effect can be modeled as optimal parameters.

3.2. PSO
PSO is a global random search optimization algorithm. PSO is used to optimize SVM learning parameters, and SVM parameter \( c \) and \( g \) are automatically determined. Assuming that the parameter sample space is \( M \)-dimensional, each sample is regarded as a particle \( i \) at a certain position in space, the moving speed of \( i \) is \( V_{id} \), the current position is \( X_{id} \), the historical optimal position is \( P_{id} \), and the optimal position of the whole population is \( P_g \). The calculation formulas for particle \( i \) moving speed and update position are equation (3) and equation (4) [7]:

\[
V_{id}^{k+1} = \omega V_{id}^k + c_1 \text{rand} \left( P_{id}^k - X_{id}^k \right) + c_2 \text{rand} \left( P_{gd} - X_{id}^k \right) \quad (3)
\]

\[
X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (4)
\]

Where, \( d = 1,2,\ldots,M; \ i = 1,2,\ldots,n; \ c_1 \) and \( c_2 \) are non-negative constants; \( \text{rand}() \) is a random number in \([0,1]\); \( \omega \) is the inertia weight, determines the impact of particles on current speed.

3.3. GS
The basic principle of GS is to divide \( c \) and \( g \) of the SVM into a grid within a given range, traverse all the points, and divide the data into groups by \( k \)-fold cross-validation, and make each subset once. The verification set, the remaining \( k-1 \) subsets are used as training sets, and the regression performance of the model is evaluated according to \( k \) models, and a set of parameters that minimize the test error is selected as the optimal parameter [8].

3.4. PSO-GS-SVM
This research uses PSO-SVM to optimize the large step parameters, delineate the optimization range, and adopts GS to refine the parameters of the range to improve the model predict performance. Figure 3 shows the forecasting process.

![Figure 3](image)

**Figure 3.** The prediction process block diagram of PV output power.

4. Example analysis
In the analysis, the measured data and related weather forecast data is from a PV power station in the summer of 2017, including direct solar irradiance, scattering irradiance, component temperature, ambient temperature, pressure, humidity, weather conditions, actual power generation, and working state of the power station. The data sampling period is 7:00-18:00, one point every 15 minutes. On sunny days, the work conditions were 26 days, of which 25 days of sample data were used as training sets, 1 day was used for prediction, and non-sunny conditions were 12 days, of which 11 days of sample data were used as training sets and 1 day was used for prediction. The model prediction performance is evaluated by using mean squared error (MSE) and squared correlation coefficient (\( R^2 \)).
After the PCA reduces the dimension, only four feature quantities are left for input. According to experience, the BP model selects two layers, in which the number of hidden layers is 9, and the training function is trainlm. For the fixed-parameter SVM model, \( c \) is 10.5 and \( g \) is 2.8. PSO-SVM sets the range of \( c \) and \( g \) to be \( 0.1 < c < 100, 0.01 < g < 1000 \), the maximum evolution maxgen is initially 200, the maximum number of populations is \( pop = 20 \), and the local search ability and global search ability parameters \( c_1 \) and \( c_2 \) are respectively 1.5 and 1.7.

Figure 4. Fitness curve MSE[PSO-SVM].

Under sunny conditions, PSO are optimized to obtain Best \( c = 11 \), \( g = 8 \), \( CVmse = 2.40 \times 10^{-4} \), and the Fitness Curve MSE is shown in Figure 4. Then GS is used to make \( c \) and \( g \) in the range \([0, 2^5]\), get Best \( c = 27 \), \( g = 27 \), \( CVmse = 1.71 \times 10^{-4} \), as shown in Figure 5. Under poor conditions, PSO are optimized to obtain Best \( c = 18 \), \( g = 3 \), \( CVmse = 2.85 \times 10^{-4} \); after GS small range optimization, Best \( c = 16 \), \( g = 5 \), \( CVmse = 2.45 \times 10^{-4} \).

Figure 6. The prediction result in sunny.

Figure 7. The prediction result in poor day.

Figure 8. Absolute error comparison chart in sunny.

Figure 9. Absolute error comparison chart in poor day.
Table 1. Performance of forecasting models.

|        | BP       | SVM      | PSO-SVM  | PSO-GS-SVM |
|--------|----------|----------|----------|------------|
| MSE    | Nice day | 1.464e-2 | 6.999e-4 | 3.044e-4   | 8.647e-5   |
|        | Poor day | 3.028e-2 | 7.644e-3 | 2.661e-3   | 1.746e-4   |
| $R^2$  | Nice day | 0.96556  | 0.99018  | 0.99526    | 0.99878    |
|        | Poor day | 0.91271  | 0.95117  | 0.97522    | 0.99898    |
| Time(s)| Nice day | 13.691   | 11.783   | 815.22     | 274.28     |
|        | Poor day | 1.81953  | 0.85646  | 207.38     | 46.821     |

Figure 6 and Figure 7 are model regression prediction curves under different working conditions. For convenience comparison, Figure 8 and Figure 9 show the absolute error curves. Table 1 lists the prediction performance indicators of each model. It can be seen that the prediction performance of all aspects of SVM is better than BP neural network; after PSO optimization, the SVM prediction index is doubled, but the time consumption is also increased. The prediction accuracy of the PSO-GA-SVM model proposed in this paper is higher than other models. The absolute error curve has the smallest fluctuation and the best fitting effect. It runs faster than the PSO-SVM model. In Figure 8, under non-clear air conditions, the light intensity is too weak before 11:00 and after 17:00, which causes the error of the measured value of the power station to become larger, so the prediction model error value is also relatively larger. However, the PSO-GS-SVM model still predicts performance better than other models.

5. Conclusion
There are many factors affecting PV output, and there are complex nonlinear relationships among various factors. The PCA reduces the dimension of each impact factor, which improving the speed of the model. The PSO-GS-SVM model is proposed to predict the photovoltaic power generation in the following day, and compares with BP, SVM and PSO-SVM prediction results. It proves that the model has the better prediction performance under different weather conditions and has strong applicability.

References
[1] Lai Changwei ,Li Jinghua ,Chen Bo ,Huang Yi Yin ,Wei Shanyang.Review of Photovoltaic Power Output Prediction Technology[J].Transactions of China Electrotechnical Society,2019,34(06):1201-1217.
[2] Shen Zhen.RESEARCH ON POWER FORECASTING OF DISTRIBUTED PHOTOVOLTAIC POWER BASED ON TRAJECTORY CHARACTERISTICS[D].Southeast University,2017.
[3] Chen Changsong,Duan Shan xu,Yin J inju n,et al.Design of Photovoltaic Array Power Forecasting Model Based on Neutral Network[J].Transaction of China Electrotechnical Society,2009,24(9):153-158.
[4] JIANG Enyu,JI Liang,XIA Nenghong,M I Yang,DENG Weihua,.Photovoltaic Power Prediction Based on Support Vector Machine[J].Journal of Shanghai University of Electric Power,2015,31(06):511-513+524.
[5] Mao Meiqin, Gong Wen jian, Zhang Liuchen, et al.Short-term photovoltaic generation forecasting based on EEMD-SVM combined method[J]. Proceedings of the CSEE, 2013, 33(34): 17-24.
[6] Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing, 14(3), pp.199-222.
[7] WANG Jidong, SONG Zhilin. RAN Ran.Short-term Photovoltaic Power Generation Rolling Forecast Based on Optimized SVM[J].Proceedings of the CSU-EPSA,2016,28(11):9-13.
[8] LI Song, FU Yun-fa, CHEN Jian, ZHANG Jian-ping.Comparative research of identification of EEG induced by Motor imagery based on five kinds of parameters optimizing SVM[J].Transducer and Microsystem Technologies,2017,36(04):42-45+50.