Research on Power Network Load Forecasting Problem Based on Machine Learning

Zhankui Wang¹, Junying Wu¹, Rui Xin¹, Tao Bai², Jianbin Zhao², Minglei Wei² and Lei Zhuang¹

¹Information and Communications Branch, Hebei Electric Power Company, State Grid, China
²Hebei Electric Power Company, State Grid, China

*Corresponding author

Abstract. Based on the existing electric power grid load forecasting method and improving the precision of electric power load forecasting, an power load forecasting model based on Support Vector Machine (SVM) and improve the parameters by Particle Swarm Optimization (PSO). Firstly, analyzed the theoretical basis of support vector machine, and the preliminary prediction model of support vector machine is established. Then, the PSO algorithm is used to iteratively choose the optimal parameters of the support vector machine parameters. Finally, the optimal load prediction model is established by the optimal parameters. The pre-processed real power data is input into the model for learning prediction, and the model prediction effect is verified by Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and prediction chart. The experimental results show that the PSO-SVM prediction model can precise forecast the power load and raised the accuracy of power load forecasting. It shows that the PSO algorithm is very effective to adjust the parameters of SVM model.

Keywords: Electric load; support vector machine; particle swarm optimization.

1. Introduction
The power load plays an extremely important role in the power management system. With the continuous development of the electric power industry and accurate prediction of the power load[1], it is possible to plan the power grid more economically and reasonably, start and stop the internal generator set, and promote power. The further development of the market and the stability and rationality of the power supply can more fully meet the needs of users.

Since the correlation between various factors and power load is not linear, there are also a large number of nonlinear factors. Therefore, it is difficult to mathematically construct corresponding explicit formulas and physical models. For this reason, mathematical statistics methods often have large errors and it is difficult to take into account many factors. For extremely large amounts of data, machine learning-based prediction algorithms will have better data processing capabilities and better extraction of useful data features.

2. Literature Review
Machine learning [2] is a method of studying how computers can imitate or realize human learning ability by algorithm, learn new knowledge and solve new problems according to their own experience and continuously improve their performance. Through machine learning, we can continuously cope with new problems and deduct and predict the characteristics and laws of future data. Vapnik and his team [3] proposed a statistical learning theory, which is a special study in the training set sample. How to
conduct theoretical science of machine learning in a limited number of cases. Statistical learning theory generally starts from some training samples, and obtains some rules that cannot be analyzed by exact theory or principle through certain methods. Based on statistical learning theory, a brand-new machine learning method Support Vector Machine (SVM) [4] is put forward. SVM is a method widely studied and applied in machine learning methods. SVM is a generalized linear classifier, which categorical data by supervised learning. This paper uses the power load forecasting model based on the SVM to preprocess the data and use it as the training set input predictive model for feature extraction. Parameter debugging, the extracted data features are tested as a training set, and the prediction set is input, and the data is predicted by the power load prediction model.

3. Support Vector Machine Algorithm and its Optimization

3.1. Support Vector Machine Algorithm

In a given training sample set \( D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}, y_i \in \{-1,1\} \), the linear classifier separates the two types of samples based on the training sample \( D \) finding a hyperplane in the two-dimensional space. As shown in Figure 1, we can easily perceive that the hyperplane represented by the green line in the figure has the strongest anti-interference ability, because the hyperplane has the largest interval from the data on both sides of the line. This means that the limitations of the training set or the noise have the greatest "tolerance" ability.

\[
\begin{align*}
&\max \frac{2}{\|w\|} s.t. \ y_i(w^T x_i + b) \geq 1, \ i = 1,2,\ldots, m. \\
&\text{Equivalent to} \ \min_{w,b} \frac{1}{2} \|w\|^2 s.t. \ y_i(w^T x_i + b) \geq 1, \ i = 1,2,\ldots, m. 
\end{align*}
\]

The above is the basic idea of the SVM model in the case of basic linear separability. The idea of the maximum interval classifier is also used in the regression model.

The SVM regression algorithm uses an \( \varepsilon - \) insensitive error function, which is defined as: if the difference \( |\hat{y}_i - y_i| \leq \varepsilon \) between the predicted value \( \hat{y}_i \) and the true value \( y_i \), no loss is considered. Otherwise, the loss cost is considered to be \( |\hat{y}_i - y_i| - \varepsilon \). That is to say, the loss function metric of the SVM regression model is:

\[
\hat{y}_i = w^T x_i + b, \ E_{\varepsilon}(\hat{y}_i - y_i) = \begin{cases} 0, & \text{if } |\hat{y}_i - y_i| < \varepsilon \\ |\hat{y}_i - y_i| - \varepsilon, & \text{else.} \end{cases}
\]
At this time, $E_{w,b}$ is used instead of the squared error term, so the minimum error function can be defined as the optimization target: 
$$\min_{w,b} C \sum_{i=1}^{N} E_{w,b}(\tilde{y}_i - y_i) + \frac{||w||^2}{2}. $$

Consider converting it into a constrained optimization problem, that is, defining two slack variables $\xi_i \geq 0$, $\tilde{\xi}_i \geq 0$ for each sample data. Rewriting the error function to a convex two optimization problem: 
$$\min_{w,b} C \sum_{i=1}^{N} E_{w,b}(\tilde{y}_i - y_i) + \frac{||w||^2}{2},$$

subject to:
$$\xi_i \geq 0, \tilde{\xi}_i \geq 0, \tilde{y}_i - (w^Tx_i + b) - \xi_i < \epsilon, (w^Tx_i + b) - \tilde{y}_i - \tilde{\xi}_i < \epsilon. \quad (5)$$

By introducing a Lagrange multiplier, the target optimization function becomes an unconstrained form:
$$L(w, b, \xi_i, \tilde{\xi}_i, \alpha, \tilde{\alpha}, \mu, \tilde{\mu}) = C \sum_{i=1}^{N}(\xi_i + \tilde{\xi}_i) + \frac{||w||^2}{2} - \sum_{i=1}^{N}(\mu_i \xi_i + \tilde{\mu}_i \tilde{\xi}_i) + \sum_{i=1}^{N} \alpha_i(\tilde{y}_i - (w^Tx_i + b) - \xi_i - \epsilon) + \sum_{i=1}^{N} \tilde{\alpha}_i((w^Tx_i + b) - \tilde{y}_i - \tilde{\xi}_i - \epsilon). \quad (6)$$

Where $\alpha > 0$, $\tilde{\alpha} > 0$, $\mu > 0$, $\tilde{\mu} > 0$, and $\epsilon$ is the coefficient of Langerang.

When the constraint is satisfied, $\max_{w,b, \alpha \geq 0} L = f$, and the Lagrangian coefficient is $0$, $\min_{w,b} \max_{\alpha \geq 0} L = \min_{w,b} f$. It can be inferred that the solution obtained by solving $\min_{w,b} \max_{\alpha \geq 0} L$ is the solution of $\min_{w,b} f$ within the constraint. First ask for partial derivatives for $w, b, \xi_i, \tilde{\xi}_i$, then bring it back to the Lagrangian function, and after simplification, we can get the function $(7)$ only about $\alpha_i, \tilde{\alpha}_i$, so maximizing the function is the next goal.

$$L(\alpha, \tilde{\alpha}) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \tilde{\alpha}_i)(\alpha_j - \tilde{\alpha}_j) \kappa(x_i, x_j) - \epsilon \sum_{i=1}^{N} (\alpha_i + \tilde{\alpha}_i) + \sum_{i=1}^{N} (\alpha_i + \tilde{\alpha}_i) \tilde{y}_i \quad (7)$$

The constraint of the formula $(7)$ is $0 \leq \alpha_i \leq C$, $0 \leq \tilde{\alpha}_i \leq C$.

The KKT optimization conditions were published independently by Karush, Kuhn and Tucker. The KKT condition is used to determine whether a solution belongs to a nonlinear optimization problem. So next we consider the KKT condition:

$$\alpha_i(\epsilon + \xi_i + w^Tx_i + b - \tilde{y}_i) = 0, \tilde{\alpha}_i(\epsilon + \tilde{\xi}_i - (w^Tx_i + b) + \tilde{y}_i) = 0, (C - \alpha_i)\xi_i = 0, (C - \tilde{\alpha}_i)\tilde{\xi}_i = 0. \quad (8)$$

It can be known from equations $(8)$ that when $\alpha_i > 0$, there must be $\epsilon + \xi_i + w^Tx_i + b - \tilde{y}_i = 0$, $\xi_i \geq 0$, and the predicted value will be smaller than the true value; When $\tilde{\alpha}_i > 0$, there must be $(\epsilon + \tilde{\xi}_i - (w^Tx_i + b) + \tilde{y}_i) = 0$, $\tilde{\xi}_i \geq 0$, then the predicted value will be greater than actual value. Moreover, for each arbitrary data point, since $\epsilon > 0$, $\alpha_i$ and $\tilde{\alpha}_i$ cannot be greater than 0 at the same time, and in order to obtain a point within the error range, there must be $\alpha_i = 0$ and $\tilde{\alpha}_i = 0$.

From the above calculation, we can know that $w = \sum_{i=1}^{N}(\alpha_i - \tilde{\alpha}_i)x_i$, and then further calculate the b value, and we can know a certain point from equation $(8)$ when the predicted value is greater than the true value, there are:

$$\xi_i = 0, (\epsilon + \xi_i + w^Tx_i + b - \tilde{y}_i) = 0$$

The derived prediction function can be derived as:
$$y(x) = w^Tx + b = \sum_{i=1}^{N}(\alpha_i - \tilde{\alpha}_i)(x_i)^T x + b = \tilde{y}_i - \epsilon - \sum_{i=1}^{N}(\alpha_i - \tilde{\alpha}_i)\kappa(x_i, x) - \sum_{i=1}^{N}(\alpha_i - \tilde{\alpha}_i)\kappa(x_i, x) \quad (9)$$

Where $(x_i, \tilde{y}_i)$ is a point where the predicted value is greater than the true value.
solution. It is a novel intelligent algorithm for group intelligence optimization. The particle swarm optimization algorithm has a simple concept, fast convergence, few parameters and no need for human adjustment, and its implementation is relatively simple.

3.2. Particle Swarm Optimization Algorithm for Support Vector Machine Parameters
The main procedures of the particle swarm optimization algorithm are:

4. Predictive Experiment

4.1. Data Sources
In order to verify the effectiveness of the constructed SVM prediction model, the power load data from June 1, 2017 to June 30, 2017 in a certain area of Hebei Province will be used. The data interval is usually three minutes, a total of 17279 sets of electricity.

4.2. Data Preprocessing
Because the data has some missing data in time series, directly reading the null value into the model will cause the model to run and the calculation error. Therefore, the null value needs to be treated as a valid value, and all missing values are set to the corresponding mean values of the upper and lower time series by time interval.

4.3. Model Evaluation Index
The Root-Mean-Square Error (RMSE) is also known as the standard error. The root mean square error is a measure of the deviation between the observed value and the true value. Calculated as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}.
\]

Mean Absolute Percentage Error (MAPE) is also a commonly used evaluation index in machine learning. It can be used to evaluate the prediction result of a model. The calculation formula is as follows:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100%.
\]

4.4. Forecast Results and Analysis
The 17279 sets of electrical load data are divided into three groups, 8000 sets of data are used as training sets to construct short-term electric load forecasting models and learn training models; 5000 sets of data are used as test sets to test the performance of post-learning electric load forecasting models; the remaining 4279 sets of data are used as predictions.
The test results of parameters without PSO optimization are as follows:
Figure 3. Load forecasting results of unoptimized parameters.
The test results of parameters optimized by PSO are as follows:

Figure 4. Load forecasting results of optimized parameters.
The evaluation indicators of the corresponding model output are as follows (Table 1, next page).

Discuss and analyze the results of Figure 3, Figure 4 and Table 1, we can find: The accuracy of the
prediction results of the PSO-SVM model is better than that of the ordinary SVM model. Therefore, the
particle swarm optimization algorithm can optimize the parameters of the SVM to calculate the power
load better, and the results can be more accurately reflected. The power load changes over a period of
time, and the anti-interference ability is stronger.

Table 1. Error analysis of prediction results.

| Predictive model | Evaluation index |
|------------------|------------------|
| SVM              | 0.2985% 0.2479%  |
| PSO-SVM          | 0.2852% 0.2256%  |

References

[1] Y. Zhang, Analysis and Research on Short-Term Forecast of Power System Load, Dissertation, Xi'an University of Technology. 2019.
[2] R. Zhang, Y. Wang, Research on Machine Learning with Algorithm and Development, Journal of Communication University of China (Science and Technology). 2016, 23(02): 10-18+24.
[3] V. Vapnik, R. Izmailov, Machine Learning, Springer. 2019, Vol. 108 (3), 381-423.
[4] G. Chu. Applications of SVM in Data Mining, Dissertation, Beijing University of Posts and Telecommunications, 2015.
[5] Q. Shao, Particle Swarm Optimization and Its Application in Engineering, Dissertation, Jilin University, 2017.
[6] N. Liu, F. Mou, Feng, Short-term Power Load Forecasting based on Least Square Support Vector Machine Optimized by NRS and PSO algorithms, Modern Electronics Technique. 2019, 42(07): 115-118+124.
[7] S. Maldonado, A. González, S. Crone, Automatic Time Series Analysis for Electric Load Forecasting via Support Vector Regression, Applied Soft Computing Journal. 2019, 83.
[8] Geng Lin, Jian Guan, Zuoyong Li, Huibin Feng. A hybrid binary particle swarm optimization with tabu search for the set-union knapsack problem[J]. Expert Systems With Applications, 2019, 135.
[9] Jinpei Liu, Piao Wang, Yanyan Huang, Peng Wu, Qin Xu, Huayou Chen. Power load combination forecasting based on triangular fuzzy discrete difference equation forecasting model and PSO-SVR[J]. Journal of Intelligent &amp; Fuzzy Systems, 2019, 36(6).
[10] Jingchao Cao, Shiqi Wang, Yan Wang, Xinfei Zhang, Sam Kwong. Content-oriented image quality assessment with multi-label SVM classifier[J]. Signal Processing: Image Communication, 2019, 78.
[11] G. M. Borkar, L. H. Patil, A novel clustering approach and adaptive SVM classifier for intrusion detection in WSN: A data mining concept, Sustainable Computing: Informatics and Systems. 2019.
[12] T. I. Dhamecha, A. Noore, R. Singh, M. Vatsa, Between-subclass piece-wise linear solutions in large scale kernel SVM learning, Pattern Recognition. 2019, 95.
[13] S. Maldonado, A. González, S. Crone, Automatic time series analysis for electric load forecasting via support vector regression, Applied Soft Computing Journal. 2019.

[14] Y. Wang, H. Zhang, G. Zhang, cPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks, Swarm and Evolutionary Computation. 2019,49.