Research on Short-Term Prediction of Power Grid Status Data Based on SVM

Jianjun Su[1], Yi Yang[1], Danfeng Yan[2], Ye Tang[2](✉), and Zongqi Mu[2]

[1] State Grid ShanDong Electric Power Research Institute, Jinan 250001, China
13953187960@163.com, yangyi814@gmail.com
[2] Beijing University of Posts and Telecommunications, Beijing 100876, China
{yandf, 2015213126}@bupt.edu.cn,
tangye_bupt@foxmail.com

Abstract. EMS (Energy management system) is a collection of computer hardware and software, which collects, monitors, controls and optimizes data provided by power control system, and provide trading scheme, security services and service analysis for power market. The prediction of status data is a basic function module of advanced application software systems. Therefore it is meaningful to do research on new method and new technology of predicting power grid status data. In this paper, support vector machine is used to do regression prediction for active power of EMS. In training process, the training set and kernel function of SVM are selected, and parameters are optimized, also, the performance of SVM is evaluated. Experiments show that SVM can get higher accuracy in short term active power prediction although the data set is small. This paper provides a new idea for related research works in electric power industry system.

Keywords: EMS · SVM · Regression prediction · Parameter optimization · Machine learning

1 Introduction

Electric power industry is one of the most important basic national industries, and is the lifeblood of the national economy, engine of economic development. It plays a crucial role for our national security, economic development, social stability, life quality. In modern society everywhere is inseparable from the power supply.

With the rapid development of power industry, research of the high precision prediction technology and application system on power state data is becoming more important and has direct and significant economic benefits and social benefits.

Active power value, usually expressed with letter *P*, is one of the most important statistic records in power grid. Active power is the power of electric energy that transfer into other forms of energy (mechanical energy, light energy, thermal energy) power, therefore the active power could reflect the usage of the whole power grid. Through active power we could have a better understanding of the power grid energy

Funded by the national high technology research and development program (863 Program) (No. 2015AA050204)

S. Wang and A. Zhou (Eds.): CollaborateCom 2016, LNICST 201, pp. 231–241, 2017.
DOI: 10.1007/978-3-319-59288-6_21

consumption, moreover we could better monitoring the operational state of the power grid. Therefore, it is very necessary to seek the way to predict the active power value in the EMS of power grid network.

The main research of this paper is to predict the active power in power condition monitoring data. We try use the existing EMS (energy management system) monitoring data, comprehensively consider meteorological factors, historical data to design and construct the input training set using support vector machine (referred to as SVM) algorithm for active power index regression prediction.

2 Research Status of Related Works

The prediction of state monitoring data of power grid is an important basic project. The improvement of the mechanism of power market will accelerate research on the new method and new technology of state power grid monitoring data prediction.

At present, there is plenty of the domestic and foreign research on the prediction of the state data of the power grid. Elke Lorenz offers a way to predict regional PV power output based on weather forecasts information up to three days ahead [1]. In order to predict the risk of failures for components and systems, Cynthia Rudin gives a general process which transforms historical electrical grid data into models by machine learning [2]. Louka P predicts wind speed of speed using Kalman filtering to give a prediction of wind power [3]. M. Carolin Mabel and E. Fernandez discuss a neural network model to predict the energy from wind farms in their paper [4].

As for using support vector machines for regression prediction, Vladimir Cherkassky and Yunqian Ma did a research on selection of parameters of support vector machines regression, and their experiments indicate under sparse sample settings, SVM regression has an excellent generalization performance for different types of additive noise [5]. Existing theory, methods, and recent developments as well as research range of SVR is discussed in Debasish Basak, Srimanta Pal and Dipak Chandra Patranabis's paper [6]. Chih-Chung Chang and CHIH-JEN LIN from offer a library for Support Vector Machines that has been developed since 2000, the library is called LIBSVM and now is widely used in machine learning and many other areas [7], which is also used in this paper.

3 The Introduction of SVM

Frequently used prediction technology includes Artificial Neural Networks [8], Times Series Analysis Method [9], Kalman Filter Analysis Method [10], Grey Models [11], Multi-output Support Vector Regression [12]. These methods have different characteristics, and they are already used in power systems. However, there are various defects, which makes practice effect not ideal. In this paper, the Support Vector Machine is applied to electric load prediction to achieve a better effect on active power prediction in the power system.

Support vector machine, namely SVM, is a new machine learning method based on Statistical Learning theory proposed by Vapnik et al. SVM is a supervised learning model used for pattern recognition, classification and regression analysis. SVM solves the problem of the linear inseparable problems by the probability of soft-margin, and introduces the kernel function to make the solution plane expand from the linear to the nonlinear [13].

To deal with linearly inseparable problems, SVM uses kernel functions. Kernel function, in essence, is a kind of mapping function, which maps the low dimensional space nonlinear problem to the high dimension space programming linear problem and then solve it.

So the basic function of a kernel function is to accept two lower dimensional space vectors, and calculate the vectors' inner-product in high dimension space after a transformation. In the nonlinear case, determine mapping function $\phi(x_i)$ is the kernel function that satisfy the Mercer conditions:

$$\phi(x_i) \cdot \phi(x_j) = K(x_i, x_j) \tag{1}$$

All functions can be used as the kernel function, as long as it satisfies the Mercer condition function. Common kernel functions include:

1. Linear Kernel Function

$$K(x_i, x_j) = x_i \cdot x_j \tag{2}$$

2. Polynomial Kernel Function

$$K(x_i, x_j) = [\gamma(x_i \cdot x_j) + c]^d \tag{3}$$

3. Radical Basis Kernel Function, which also called Gauss Kernel Function, the expression is:

$$K(x_i, x_j) = exp[-\frac{|x_i - x_j|^2}{\sigma^2}] \tag{4}$$

Where σ is the Radial Basis Radius, take $g = \frac{1}{\sigma^2}$ into formula (4) is another common expression of Gauss's function:

$$K(x_i, x_j) = exp(-g|x_i - x_j|^2) \tag{5}$$

4. Sigmoid Kernel Function

$$K(x_i, x_j) = tanh(\gamma(x_i, x_j) + c) \tag{6}$$

4 Scenario Description and Definition of the Research Problem

4.1 Data Preprocessing and Scenario Description

Main source of data for the research is from the history EMS state information of an electricity substation of the Shandong province power grid in June 2015, including meteorological temperature information. Table 1 are examples of the contents of the meteorological table of the EMS records.

Table 1. Examples of meteorological data in the EMS

ID	Device	Time	Temperature	Humility
1	StationA	2015-07-15 00:00:00	23.2	099
2	StationA	2015-07-15 00:10:00	23.5	098

Table 2 shows the records of transformer equipment state data in the EMS. It can be seen in the table recording the record ID, equipment ID, the site code, as well as the record time, active power value orderly.

Table 2. Examples of equipment state data in the EMS

ID	dev_id	dev_name	Time	*P*
1	1800002459	Station C	2015-07-15 00:00:00	147.7919
2	1800002459	Station C	2015-07-15 00:05:00	150.3291
3	1800002459	Station C	2015-07-15 00:10:00	145.8889

Data pre-processing can be divided into two steps, data cleaning, and data association to integrate into the SVM training data set. After remove redundant, abnormal and wrong information, there are 4018 records of temperature data, also 4018 records of EMS's active power value data.

By simple statistics, we can see the temperature in three days shown in Fig. 1.

The active power of the corresponding time range is shown in Fig. 2.

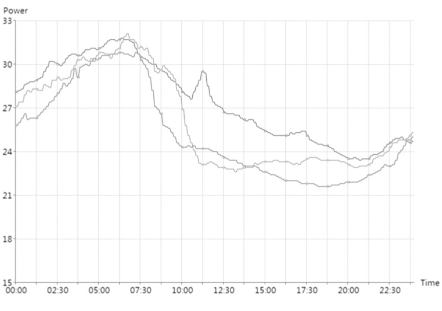

Fig. 1. Range of temperature within 3 days

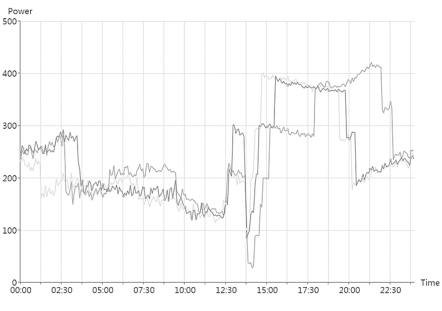

Fig. 2. Range of active power within 3 days

It can be clearly seen from the two pictures, the temperature and active power show a relatively fixed variation law. This rule can be used as a reference for the design and construction of the training set of data in the following chapters.

In this scenario, we hope to develop a good regression prediction model, making grid equipment meteorological data, environment data, state information as input to predict the specific value of active power(referred to as *P*) in the future time. The prediction can be a reference to the assessment of whole grid operation state, avoiding major accidents, making decisions of power grid operation and so on. Due to the short term prediction's characteristics such as efficient and agile, as well as to be able to get accurate results using less sample data, short term prediction is very suitable for the support vector machine learning methods for regression prediction which has a small demand on data size.

4.2 Problem Definition

There is n set of records of active power (*P*) in future time corresponding to input samples are abstracted as n dependent variables, and are denoted by a vector:

$$Y_n = (y_1, y_2, \ldots, y_n) \tag{7}$$

In formula (7), n is the number of samples, and y_i indicates the active power (P) of the predicted time of the ith input sample P. Independent variable is:

$$X_n = (x_1, x_2, \ldots, x_n) \tag{8}$$

And x_n is a $n*N$ matrix. The ith row of the matrix x_i is a vector that comprises N dimensional variables corresponding to y_i where N dimension representing N kinds of grid environment factors (temperature, historical data, etc.), and N can be one or more of these factors. Moreover n is the number of samples, a total of n samples is independent variables. For the specific selection and design of training data set x_n will be discussed in the following chapters.

Do regression training using model *M*, which takes x_n as input vector, and y_n as SVM's label.

In regression prediction, using the trained model *M* to predict data vector x_s:

$$X_s = (x_1, x_2, \ldots, x_s) \tag{9}$$

As input, X_s is a $n*S$ matrix that has the same structure with input vector X_n. The output is Y_s:

$$Y_s = (y_1, y_2, \ldots, y_s) \tag{10}$$

Each value y_i in Y_s represents the corresponding input vector in X_s, which is the active power value corresponding to the ith row x_i.

Moreover, to measure the error and analyze the result, in this paper, the mean relative error e_{MRE} and root mean square error e_{RMSE} are used as the basis for judging the effect of various prediction methods. Their calculation methods are as follows:

$$e_{MRE} = \frac{1}{N}\sum\nolimits_{i=1}^{N} |\frac{L(i) - L'(i)}{L(i)}| \times 100\% \quad (11)$$

$$e_{RMRE} = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N} (\frac{L(i) - L'(i)}{L(i)})^2} \times 100\% \quad (12)$$

In the formulas, $L(i)$ representing the actual active power value at a certain time and $L'(i)$ representing the predicted active power value.

At the same time for every moment of the actual active power value and prediction active power value, we make the e_{single} as a single moment prediction results of the error percentage, the formula is as follows:

$$e_{single} = |\frac{L(i) - L'(i)}{L(i)}| \times 100\% \quad (13)$$

In this paper, we make 5% as the judging criterion, if a result having $e_{single} > 5\%$, we consider this prediction result fails. We define the qualified rate of r for active power prediction of algorithm result:

$$r = \frac{p'}{p} \times 100\% \quad (14)$$

Where P' is the number of results satisfy $e_{single} <= 5\%$, that is the number of qualified prediction results. P is the total number of prediction results.

In summary, we select the prediction results pass rate r, the mean relative error e_{MRE} and the root mean square error e_{RMSE} as the gauge of prediction accuracy rate.

5 Experiment Process and Result of the Algorithm

5.1 Brief Introduction of the Algorithm

The algorithm of this paper is divided into four parts, the first part carries on the construction of the training set data in different ways, and compares the results to select the optimal design, the second part is the selection of the kernel function, the third part is the adjustment of penalty factor and kernel function under the condition of the selected kernel function, the fourth part is the analysis of experimental results.

5.2 The Selection of Training Set

The randomness of the power measurement index is very strong and has many influence factors, so the short-term active power prediction is a multi-variable regression prediction problem.

As described in the previous chapter, the active power y_i of the predicted time point is the output value of the function. And the factors that affect the y_i, such as: historical data, temperature and meteorological information, as the input vector x_i of function. So we take multiple designs on input vector x_i checking the effect of prediction based on SVM regression, specific designs are as follows.

Scheme 1 is designed as follows:

$$x_i = \{b_1, b_2, b_3, b_4\} \tag{15}$$

In the formula b_1, b_2, b_3, b_4 represent the active power values of the 4 records before the target prediction time.

Scheme 2 is designed as follows:

$$x_i = \{t_1, b_1, b_2, b_3, b_4\} \tag{16}$$

In the formula b_1, b_2, b_3, b_4 represent the active power values of the 4 records before the target prediction time. And t_1 represents the temperature data of the prediction time most recently

Scheme 3 is designed as follows:

$$x_i = \{h_1, h_2, b_1, b_2, b_3, b_4\} \tag{17}$$

In the formula b_1, b_2, b_3, b_4 represent the active power values of the 4 records before the target prediction time. And h_1 represents the active power value of yesterday at the same time with the prediction time, h_2 represents the active power value of last week at the same time with the prediction time.

Scheme 4 is designed as follows:

$$x_i = \{t_1, h_1, h_2, b_1, b_2, b_3, b_4\} \tag{18}$$

In the formula b_1, b_2, b_3, b_4 represent the active power values of the 4 records before the target prediction time. And t_1 represents the temperature data of the prediction time most recently. And h_1 represents the active power value of yesterday at the same time with the prediction time,. h_2 represents the active power value of last week at the same time with the prediction time.

For the four design schemes of the input vector, we select 5 consecutive days that have more than 1440 records to structure training data set. And the default Gauss kernel function and the unmodified standard parameters are used as the configuration of the algorithm model. Predict the active power value of 288 time points in the next 1 day.

The assessment of the results is as described in the previous. The results are shown in the following table (Table 3):

Table 3. Experimental results of different training set

Design scheme	r	e_{MRE}	e_{RMSE}
Scheme 1	66.315%	8.079%	16.055%
Scheme 2	68.571%	6.422%	12.158%
Scheme 3	68.214%	6.386%	12.066%
Scheme 4	68.571%	6.439%	12.062%

Through the experimental results on the table, the difference between the results of the four designs is small. Comparing the results of four designs, we choose scheme 4 as the training set design.

5.3 The Selection of Kernel Function

As the formula (1) mentioned in the third chapter, the SVM kernel function is required to meet the Mercer condition. Four commonly used kernel functions introduced before have different characteristics:

Linear kernel function is mainly used in the case of linear separable. Less parameters, faster speed, for the common data, the classification effect can achieve an ideal result.

Polynomial kernel function in SVM is not commonly used, much more used for the NLP Natural Language Processing.

Radical Basis Kernel Function. Also known as the Gauss kernel function, which is mainly used in the case of linear non separable. The number of parameters is relatively more, and the classification results are very dependent on the parameters.

Sigmoid kernel function. When choose sigmoid function as kernel function, support vector machine is a multilayer perceptron neural network.

In practical application, we can choose the reasonable kernel function according to the scope of the application of the four kinds of kernel functions. In this paper, four kinds of kernel functions are compared with the standard parameters, and the results are shown in Table 4:

Table 4. Experimental results of different kernel function

Kernel function	r	e_{MRE}	e_{RMSE}
Linear kernel function	58.947%	9.251%	17.601%
Polynomial kernel function	6.667%	26.846%	32.199%
Radical basis kernel function	68.214%	6.386%	12.066%
Sigmoid kernel function	13.684%	14.148%	17.247%

According to the experimental results, the performance of the radial basis kernel function is better. So it is more appropriate to select the radial basis kernel function (i.e. Gauss's function) as the kernel function the SVM model.

5.4 Adjustment of Penalty Factor and Parameters of Kernel Function

For the parameter adjustment of Gauss kernel function, we mainly adjust the parameters of the Gauss kernel function *C* and gamma.

In summary, gamma is the balance factor between the sparse degree of result expression and the density of data points. Penalty factor *C* is to prevent the model from giving up some important data in the learning process and to avoid loss of data.

We implement the optimization and adjustment to parameter of the kernel function, the specific steps are as follows:

1. Set range of parameter gamma, parameter *C*, set the search step size.
2. For each pair of gamma, *C*, calculate SVM's pass rate of *r* for the test data;
3. Form each pair *C*, gamma, choose the optimal *C*, gamma value by the rate r.
4. If pass rate *r* cannot achieve the requirements, select a smaller search range based on current results, reduce the search step distance for search.

In this paper, the initial selection of *C* search range is $[2^{-8}, 2^{8}]$, step distance is 0.5; Gamma search range is [2–10, 23], step distance is 0.5. The algorithm is trained by libSVM. Training data is 48 consecutive hours of the historical data from a substation in Shandong province since June 9, 2015, including weather condition. The total number of the training data is 573. And 141 pieces of test data of the next 12 h of data is used for regression prediction.

The results of some parameters are shown in the following Table 5.

Table 5. Partial results of some parameters of Gauss kernel function

C	gamma	*r*
1	1	25.532%
1	0.1	25.532%
1	0.01	25.532%
1	0.0001	22.695%
16	0.1	25.532%
16	0.01	25.532%
16	0.0001	29.087%
32	0.1	19.149%
32	0.01	24.531%
32	0.0001	46.099%
32	0.000001	80.142%
64	0.0001	60.284%
64	0.00001	72.340%
64	0.000001	80.851%
100	0.00001	71.631%
100	0.000001	82.269%
128	0.0001	58.156%
128	0.000001	80.851%

Through the results of the test, the pass rate r achieved the highest value of 82.269% when C = 100, Gamma = 0.000001. So in this paper, we choose C = 100, Gamma = 0.000001 as the optimal parameter of the algorithm model.

Compared with the Sect. 5.3 default SVM regression prediction model using the Gauss kernel function, the prediction results are shown to the following Table 6.

Table 6. The comparison of results before and after adjust the parameters of kernel function

SVM model	r	e_{MRE}	e_{RMSE}
Default Gauss kernel function	68.214%	6.386%	12.066%
After parameter optimization	82.269%	3.451%	6.419%

We can see through parameter optimization, algorithm model's pass rate r, mean relative error e_{MRE} and root mean square relative error e_{RMSE} gained a substantial increase, and we get better the regression prediction results.

5.5 Analysis of Experiment Result

The experimental process of the whole algorithm is the design and implementation of the algorithm, the selection of the training set, the selection of kernel function and the optimization of the parameters, which greatly improves the accuracy of prediction.

According to the results of the previous experimental process, we select 48 h substation EMS data since June 9, 2015 as training data, and according to Sect. 5.2 we choose scheme 4 as the structure of the training set, and as the comparison result of Sect. 5.3 we use Gaussian kernel function. Meanwhile we set the parameters at Sect. 5.4. Then we predict the active power value of the next 12 h using the algorithm we developed. The comparison between the predicted data and the real data of the active power value is as follows (Fig. 3):

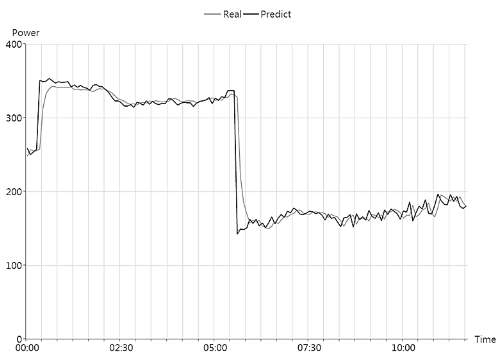

Fig. 3. Comparison between the predicted value and the real value

We can see through the selection of kernel function and parameters optimization, algorithm performs well in the prediction, the rate of accuracy is relatively high.

6 Summary

Short-term prediction of the power system's state data based on SVM algorithm model can achieve a high level of regression prediction accuracy. To a certain extent, in the case of fewer samples, SVM is a relatively good regression prediction algorithm model. Through the selection of kernel function and kernel function parameter optimization, SVM's performance is more excellent, which proved the necessity and importance of SVM parameter optimization. This paper provides guide for the practice of prediction of power grid status data.

References

1. Lorenz, E., Hurka, J., Heinemann, D., et al.: Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J. Sel. Topics Appl. Earth Observations Remote Sens. **2**(1), 2–10 (2009)
2. Rudin, C., Waltz, D., Anderson, R.N., et al.: Machine learning for the New York City power grid. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 328–345 (2012)
3. Louka, P., Galanis, G., Siebert, N., et al.: Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J. Wind Eng. Ind. Aerodyn. **96**(12), 2348–2362 (2008)
4. Mabel, M.C., Fernandez, E.: Analysis of wind power generation and prediction using ANN: A case study. Renew. Energy **33**(5), 986–992 (2008)
5. Cherkassky, V., Ma, Y.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw. **17**(1), 113–126 (2004)
6. Basak, D., Pal, S., Patranabis, D.C.: Support vector regression. Neural Inf. Process. Lett. Rev. **11**(10), 203–224 (2007)
7. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) **2**(3), 27 (2011)
8. Kara, Y., Boyacioglu, M.A., Baykan, Ö.K.: Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange. Expert Syst. Appl. **38**(5), 5311–5319 (2011)
9. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., et al.: Time Series Analysis: Forecasting and Control. Wiley, New York (2015)
10. Wei, G., Ling, Y., Guo, B., et al.: Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman filter. Comput. Commun. **34**(6), 793–802 (2011)
11. Hamzacebi, C., Es, H.A.: Forecasting the annual electricity consumption of Turkey using an optimized grey model. Energy **70**, 165–171 (2014)
12. Wang, S., Hsu, C.H., Liang, Z., et al.: Multi-user web service selection based on multi-QoS prediction. Inf. Syst. Front. **16**(1), 143–152 (2014)
13. Maji, S., Berg, A.C., Malik, J.: Efficient classification for additive kernel SVMs. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 66–77 (2013)