An electric power generation forecasting method using support vector machine

Li Guo, Jinhao Chen, Fukui Wu and Manran Wang

College of Information Engineering, Hubei University for Nationalities, Enshi, People’s Republic of China

ABSTRACT
The diversity of power generation provides an important guarantee for the electric reliability of human society. The forecasting of power generation is an important topic in the electrical industry. However, most of recent work are focus on some special type power generation, overall electric load forecasting is lacking attention. In order to improve practical applications, this paper proposes a power generation predication method based on one of popular machine learning algorithm that is support vector machine, so as to predict both overall power generation and some special types of power generation. The nonlinear relation of electric net power generation is explored by historical monthly recorded data, this relation can help the predication of net electric generation for the next month. Experimental results show that our proposed electric generation forecasting method based on support vector machine can get suitable predication model and achieve high predicted precision, which is in accordance with the real data in the record.

1. Introduction

Electricity is a vital factor in the development for the national economy. One of the characteristics of electric power production is that power cannot be stored in large quantities, the power consumption and the power generation should be balanced in real time. If it is not satisfied, the voltage quality will be severely affected. The high voltage fluctuation not only increases the power consumption but also affects the life and safety of the electrical equipment. In order to realize the balance between power generation and electricity consumption, the prediction of electric net generation is very important for the national economy. It provides great significance to reasonably allocate the power generation material and make an optimal power planning decision by the power department of a country. Electric net power generation is equal to net power multiplied by running hours. The net power is also called the sum of power supply load, which is equal to the sum of the network loss and the integrated electrical load.

Many researchers have proposed some methods to predict the amount of electricity in their work. A. Bugala presented a classical statistical method based neural modelling that has been conducted with the use of Statistica software to obtain short-term forecasting of electric energy from photovoltaic conversion (Bugala et al., 2018). Zhong et al. integrated particle swarm optimization algorithm and gray theory model to improve the prediction accuracy from a large amount of real data of two separate power stations in China (Zhong, Yang, Cao, & Yan, 2017). Kim et al. proposed a daily prediction model based on the weather forecast information for solar power generation, which is embedded into a solar photovoltaic (PV) monitoring system that is commercially used in Korea, and it is shown to perform better than the existing prediction models (Kim, Kim, Yoo, & Lee, 2017). E. Gonzalez proposed an autoregressive integrated moving average model (ARIMA) to efficiently predict monthly trend and fluctuation electric energy (Gonzalez-Romera, Jaramillo-Moran, & Carmona-Fernandez, 2006). There were some works using other methods to forecast power generation (Eseye, Zhang, & Zheng, 2018; Leone, Pietrini, & Giovannelli, 2015; Mohandes, 2002; Niu, Wang, & Wu, 2010; Xia, Zhang, & Cao, 2017; Zeng & Qiao, 2013; Zhirui & Runing, 2013). Türkay used the real electrical load values, temperature and electricity price as economic factors from 2006 to 2009, so as to predict daily peak load values in April 2010 (Türkay & Demren, 2012). Though there are many methods for power generation prediction, the including prediction factor is different in various situations, so it is not acceptable in general application. In order to extend the power generation prediction to a wider application, this paper proposed an electricity power generation forecasting method using support vector machine (SVM), which modelled the nonlinear relations with the generating capacity of coal, oil, natural
gas, nuclear electric power, and some other conventional hydroelectric power to analyse the influence of raw materials on electrical net power generation, so as to get accurate experimental forecasting result.

This paper is organized as follows. In Section 2, we discussed the principles and applications of other common methods. Section 3 introduced the principles and steps of our own method. Then in Sections 4 and 5 show the experimental results and draw a conclusion.

2. Support vector machine

Support vector machine (SVM) is a method to achieve the minimum empirical risk and confidence range by seeking the minimum structured risk (Zhirui & Runing, 2013). The main idea of SVM in regression is that the input must be mapped into a higher dimensional feature space with the function $\phi(x)$ (Türkay & Demren, 2012), here $\phi(x)$ is a nonlinear mapping from the input space to the high dimensional feature space. The predication result of SVM is generalized by defining the insensitive loss function $\varepsilon$, making it suitable for function regression (Mrabe, 2017). SVM can be divided into two types, one is linear separation and the other is nonlinear separation, which is detailed as follows.

2.1. Linear separation

In SVM, the object is to look for an optimal division line, which can easily make the largest margin on both sides. In this case, several data points at the edge are called support vector. Assuming that the data is linearly separable, the two kinds of data can be separated with a straight line, which is equivalent to a hyperplane. The $y$ of the data points on the hyperplane side is all $-1$, and the corresponding $y$ on the other side is $1$. This hyperplane can be represented by the classification function $y = f(x) = wx + b$. When $f(x)$ is equal to 0, $x$ is the point on the hyperplane, and the point where $f(x)$ is greater than 0 corresponds to the point of $y = 1$, and the point where $f(x)$ is less than 0 corresponds to the point of $y = -1$, as shown in Figure 1. Here $x$ is a sample point, $w$ is a vector perpendicular to the hyperplane, $||w||$ represents the norm, $b$ is a constant.

2.2. Nonlinear separation

In the nonlinear situation, the key point is how to select a suitable kernel function $K$, so as to solve linear and nonlinear separation in the original space by mapping the data to the high dimensional space. The support vector machine performs the calculation in the low dimensional space firstly, and then maps the input space to the high dimensional feature space with kernel function, and finally constructs the optimal separation hyperplane in the high dimensional feature space. Some commonly used kernel functions are introduced as follows.

a. Linear function: $K(x, x_i) = (x, x_i)$;

b. Polynomial function: $K(x, x_i) = (s(x, x_i) + c)^d$, here $s$ and $d$ are the parameters of the above formula;

c. Radial basis function (RBF): $K(X, X') = e^{\frac{||x - x'||^2}{2\sigma^2}}$. Here $\sigma$ is the related parameters;

d. Sigmoid: $K(x, x_i) = \tanh(v(x \cdot x_i) + c)$. Here $v$ and $c$ are related parameters.

In this paper, above different kernel function is adopted respectively for better selection for prediction of total net electricity power generation shown in Fig. 2(a). The experimental results are given as follows: the correlation coefficient $R$ obtained by linear kernel function is 89.92%, 92.35% is obtained by polynomial kernel function, 93.05% is obtained by, and 52.11% is obtained by the sigmoid kernel function. It is clearly shown that when the kernel is radial basis function can get the most accurate prediction, so it is selected in the following experiments.

3. Power generation prediction method based SVM

In this paper, the SVM algorithm is implemented in the LibSVM 3.22, which is proposed by Prof. Lin Zhiren (Chung & Lin, 2011). In order to further prove the efficiency of SVM, the forecasting result by Elman neural network also provided in the experimental results for better comparison.
Figure 2. Electricity net power generation of all material sources: (a) total net electricity power generation, (b) electricity net power generation from coal, (c) electricity net power generation from petroleum, (d) electricity net generation from natural gas, (e) electricity net generation from nuclear electric power, (f) electricity net generation from conventional hydroelectric power.

3.1. Construction of predictive factors

Different power generation sources produce different power generation data, the power generation is a closely related factor for the total power generation. These factors should be considered in the prediction of the total net generating capacity, and a reasonable way of the power grid operation to improve the efficiency, security and stability of the power grid. In this section, different power generation sources adopted in our experiments including from January 1973 to October 2017 for each month’s electricity net generation, which are clearly shown in Figure 2. These data come from an available
recorded dataset of U.S. Energy Information Administration (https://www.eia.gov/electricity/data.php), it including 538 rows and 6 columns in the official website dataset. (The 538 line represents 538 months from January 1973 to October 2017, The six columns represent six different power generation materials that include coal, petroleum, natural gas, nuclear electric power, conventional hydroelectric power and the sum of them.)

In Figure 2, the horizontal axis represents all of the months from January 1973 to October 2017, and the vertical axis represents electricity net power generation of all different sources respectively. Figure 2(a) describes the total electricity net power generation, we can see that it is increased annually. Figure 2(b) is the produced electricity net generation from coal, which is an important electricity source in many countries all over the world, it also has a vital influence in the total power generation. It is shown that the maximum power generation is the 427th months, which almost 1.9*10^15 Million Kilowatthours. Figure 2(c) is described electricity generation caused by petroleum, it almost has the same important effect as coal, it is an inevitable power generation source in the total power generation capacity. Figure 2(d) is a kind of clean and environmental protection source, which is widely used in power generation and modern industry, that is natural gas. It also has great research significance for power generation forecasting. Figure 2(e) is a kind of efficient and economical clean energy source nuclear power, it has been paid more attention in the field of power generation. Figure 2(f) shows hydroelectric power generation source which is widely used in our country. It can be seen from the diagram, hydroelectric pumped storage is most sensitive to seasonal influence (Arif & Zaman, 2017).

3.2. Prediction model construction

3.2.1. Model constriction

The SVM training set consists of six inputs, namely, monthly electricity net generation of coal, petroleum, natural gas, nuclear electric power, conventional hydroelectric power and the sum of them. That is to say, the total electricity output of next month is determined by the following factors, which were total coal power generation, total oil power generation, total natural gas power generation, total nuclear power generation and total conventional water power generation, all of them are independent variables of the total net electricity generation of each month. The total net generating capacity of each month is the dependent variable.

The flow of the proposed method is shown in Figure 3.

3.2.2. Data preprocessing

The normalization of the training set and the prediction set is processed, and the normalized mapping is treated as follows (Meena1, Panwar, Negi, & Jariai, 2012; Tasdighi & Kezunovic, 2017):

\[ f : x \rightarrow y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} + 1 \]

The raw data is normalized to the [1, 2] range. The singular sample data refers to the sample vectors which are especially large or very small compared with other input samples. The existence of singular sample data will cause an increase in training time and may also lead to non-convergence. Therefore, it is necessary to normalize the preprocessed data before training when there is a singular sample data. The result of this normalization is shown in Figure 4.

Figure 4 describes the total electricity net power generation, the normalized diagram compared with Figure 2(a), all departments refer to all power generation departments in U.S. Energy Information Administration provided on the official website.

3.2.3. Parameter selection

In this paper, the grid search method is adopted to obtain the optimal result. This method including two important parameters \( c \) and \( g \). Here \( c \) represents as penalty parameter, adjusting the weights of preferences in two indicators (interval size, classification accuracy) in the optimization direction. The higher \( c \) means the lower error. Parameter \( g \) represents gamma setting in the kernel function, the higher \( g \) means minor of the support vectors, which will affect the training and prediction speed. In this part, we will discuss the selection of these important parameters.

Rough choice: firstly, \( c \) is selected as \( 10^{-8} - 10^{-1} \), \( g \) is selected as \( 10^{-8} - 10^{-8} \) by experience. The range of rough selection is set according to experience, and the
function is to roughly determine the range of $c$ and $g$, providing a set range assistance for fine selection.

Fine selection: after rough choice, $c$ is selected as $10^{-4}–10^{+4}$, and $g$ is selected as $10^{-4}–10^{+4}$ in fine selection. The step length of $c$ and $g$ is set to 0.05 in this process.

In Figure 5, the best cross-validation MSE $= 0.007661$ is achieved after the optimal process, the best parameter tuning result is $c = 1$ and $g = 12.996$ in this experiment.

3.3. Overall and other types of power generation

Based on Section 3.2 referred prediction model, this part aims to predict the amount of electricity generated from different power generation resources. It can be seen from the error and the intensive reading of each experiment, and verify that the experiment has the characteristics of

the wide application and high precision. The analysis of the results obtained by different processing of the same data is of great significance for verifying the correctness of an experimental result.

Figure 6 shows a comparison between the regression prediction data and the original data of the year January 1973 to October 2017. Figure 6(a) predicts the amount of electricity net power generated from coal, the test results show that mean square error $\text{MSE} = 0.00587014$ and correlation coefficient $R = 90.8116\%$. Figure 6(b) predicts the amount of electricity generated from petroleum, the test results show that mean square error $\text{MSE} = 0.00290585$ and correlation coefficient $R = 92.5474\%$. Figure 6(c) predicts the amount of electricity generated from natural gas, the test results show that mean square error $\text{MSE} = 0.00227314$ and correlation coefficient $R = 94.9718\%$. Figure 6(d) predicts the amount of electricity generated from nuclear electric power, the test results
show that mean square error $\text{MSE} = 0.0024255$ and correlation coefficient $R = 97.1019\%$. Figure 6(e) predicts the amount of electricity generated from conventional hydroelectric power, the test results show that mean square error $\text{MSE} = 0.008222$ and correlation coefficient $R = 80.9229\%$.

4. Experimental results and analysis

4.1. Comparable experiment of various independent variables

On the basis of the original provided experimental data by U.S. Energy Information Administration, 538 rows and 6 columns are selected in this experiment. The total of electricity net generation is shown in Figure 7.

Figure 7 shows prediction model referred in Section 3.2, 538 rows and 6 columns sample data were selected from the original experimental data to predict the total electricity net power generation. This figure detailed describes the curves of the raw data and the regression forecasting data. In order to further clearly show the comparable result of raw data and regression forecast data in the left part in Figure 7, 50 groups of data (from September 1989 to November 1993) are selected to describe the detailed difference which is shown in the right of Figure 7. Here mean square error (MSE) is adopted to evaluate

Figure 6. Prediction comparison of different data source: (a) regression curves of electricity net power generation from coal, (b) regression curves of electricity net power generation from petroleum, (c) regression curves of electricity net power generation from natural gas, (d) regression curves of electricity net power generation from nuclear electric power, (e) regression curves of electricity net power generation from conventional hydroelectric power.
the prediction difference of forecasting value and true providing value, which value is $\text{MSE} = 0.00235087$ and correlation coefficient $R = 96.218\%$.

4.2. Experimental results analysis

Based on Section 3.2 proposed prediction model, the amount of generation electricity prediction is performed in this part. The experiment selects three input variables, five input variables, six input variables and nine input variables respectively for comparison.

In Table 1, three types of electricity generation are included here. The second row represents different types of electricity generation, which including the total generation, generation from coal and generation from nuclear power. In the total electricity net generation, 538 represents 538 months data (from January 1973 to October 2017), number 3, 5, 6 and 9 represents the different size of input variable (e.g. number 5 represents electricity generation from coal, petroleum, nuclear electric power, conventional hydroelectric power of all sectors, etc., except for those sources, there are some other generation sources, such as waste, geothermal and so on).

Figure 6. Continued.

Figure 7. Comparison of the raw data and regression forecast data.
Table 1. Comparable forecasting result of SVM with different size variables.

| Different types of electricity generation | Correlation coefficient $R$ |
|------------------------------------------|-----------------------------|
| The total of electricity net generation  |                             |
| 538*3                                    | 91.29%                      |
| 538*5                                    | 96.92%                      |
| 538*6                                    | 96.22%                      |
| 538*9                                    | **97.0056%**                |
| Electricity net generation from coal     |                             |
| 100*5                                    | 68.09%                      |
| 300*5                                    | 75.32%                      |
| 400*5                                    | 79.22%                      |
| 538*5                                    | **90.1372%**                |
| Electricity net generation from nuclear power |                     |
| 100*5                                    | 20.21%                      |
| 300*5                                    | 60.26%                      |
| 400*5                                    | 87.16%                      |
| 538*5                                    | **96.4311%**                |

Figure 8. Comparable result of a relative error.

The third and fourth row represents electricity generation from a single source that is coal or nuclear power. Similar to the former explanation, 5 represents five features, which is composed of electricity net generation from coal of all sectors, coal consumed by the electric power sector, coal stocks of the total, coal stocks of the electric power sector, and electricity net generation total of all sectors. Number 100, 300, 400 and 500 represents 100 groups data (from July 2009 to October 2017), 300 groups data (from December 1992 to October 2017) 400 groups data (from July 1984 to October 2017), and 538 groups data (from January 1973 to October 2017) respectively. Experimental results shown in Table 1 reflects that the more features cause higher correlation coefficient $R$, that is more accurate forecasting result.

4.3. Comparable experiment with Elman algorithm

In all kinds of neural networks, the forward neural network is suitable for the prediction of net power generation for its input delay. According to the historical data of net electric generation, we select the input and output nodes of feedback neural network to reflect the inherent rule of net electric generation. so as to predict the purpose of generating electricity in the future period (Darong, Jianping, & Ling, 2015; Huang, Chen, Sun, & Zhao, 2017).

In the experiment, the net electric generation in the first 537 months is used as the training sample of the network (which fuel type is total, coal and nuclear power), so the size is 537*3. Each three months data are combined to construct a vector as input data, this allows us to get 534*9 sets of training samples. Then, we set the 538th-month data as the test sample of the network, so as to verify the prediction efficiency of the power electricity in the current month, which is shown in Figure 8. In this experiment, the factor is set as follows: the number of neurons in hidden layer is 4, the value is 7, 11, 14 and 18 respectively. The threshold is set to be $[0, 0, 1, ..., 0, 1]$, all of that is 9 pairs.

In Figure 8, the horizontal axis represents different fuel type (data source), in which 1 represents electricity net power generation total of all sectors, 2 represents electricity net power generation from coal of all sectors, and 3 represents electricity net power generation from nuclear power of all sectors. We chose the number of neurons to be 7 in this experiment.

In order to further show the superiority of our proposed method in power generation forecasting, we also provide the comparable experimental results of SVM and Elman neural network algorithm in Table 2. We provide four groups of experimental result, which is the sample size of 100*3, 260*3, 400*3 and 538*3 respectively. Similar to the above experiments, the row represents power generation data chosen from these months respectively, and the column represents the different fuel type (3 represents the data chosen from the total, coal and hydroelectric pumped storage). In Table 2, we can see that the forecasting result of SVM is obviously superiority than Elman neural networks algorithm.

Table 2. Comparison SVM with Elman of experimental results.

| Experimental type | Relative error |
|-------------------|----------------|
|                   |                |
| SVM               | 0.011089       |
| 100*3             | 0.009198       |
| 260*3             | 0.012662       |
| 400*3             | -0.007856      |
| 538*3             | **0.04331362** |
| Elman             | 0.03785972     |
| 100*3             | 0.0536324      |
| 260*3             | **0.08417252** |
| 400*3             |                |
| 538*3             |                |

5. Conclusion

From January 1973 to October 2017, the total electrical net power generation, coal generation, petroleum generation, natural gas generation, nuclear electric power generation, conventional hydroelectric power generation are used to predict net power generation of each sector next month by SVM in this paper. Compared with previous
work in this field, our work can achieve more accurate prediction result compared with real raw data in the record. In the future work, we will explore the advantage of less independent variables and simple calculation of the advanced SVM or other superior forecasting method.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by Natural Science Foundation of Hubei Province: [grant number 2014CBF612, 2015CFC781]; National Natural Science Foundation of China: [grant number 61663008]; Ph.D. Technology Program: [grant number MY2014B018].

**References**

Arif, M., & Zaman, U. (2017). Power grid classification through electrical network frequency. *Bangladesh University of Engineering and Technology*, 1–43.

Bugała, A., Zaborowicz, M., Boniecki, P., Janczak, D., Koszela, K., Czekała, W., et al. (2018). Short-term forecast of generation of electric energy in photovoltaic systems. *Renewable & Sustainable Energy Reviews, 81*, 306–312.

Chung, C., & Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 1–27.

Darong, H., Jianping, T., & Ling, Z. (2015). A fault diagnosis method of power systems based on gray system theory. *Mathematical Problems in Engineering, 3*, 1–11.

Eseye, A. T., Zhang, J., & Zheng, D. (2018). Short-term photovoltaic solar power forecasting using a hybrid wavelet-pso-svm model based on scada and meteorological information. *Renewable Energy, 118*, 357–367.

Gonzalez-Romera, E., Jaramillo-Moran, M. A., & Carmona-Fernandez, D. (2006). Monthly electric energy demand forecasting based on trend extraction. *IEEE Transactions on Power Systems, 21*(4), 1946–1953. [https://www.eia.gov/electricity/data.php](https://www.eia.gov/electricity/data.php)

Huang, D., Chen, C., Sun, G., & Zhao, L. (2017). Linear discriminant analysis and back propagation neural network cooperative diagnosis method for multiple faults of complex equipment bearings. *Acta Armamentarii, 38*(3), 1649–1657.

Kim, J., Kim, D., Yoo, W., & Lee, J. (2017). Daily prediction of solar power generation based on weather forecast information in Korea. *IET Renewable Power Generation, 11*(10), 1268–1273.

Leone, R. D., Pietrini, M., & Giovannelli, A. (2015). Photovoltaic energy production forecast using support vector regression. *Neural Computing & Applications, 26*(8), 1955–1962.

Meena1, V., Panwar, R., Negi, A., & Jaria1, R. (2012). Performance analysis of an induction motor drive by SVM-direct torque control method using MATLAB simulation. *International Journal of Electronics, Electrical and Computational System, 6*, 12–16.

Mohandes, M. (2002). Support vector machine for short-term electrical load forecasting. *International Journal of Energy Research, 26*(26), 335–345.

Mrabe, H. (2017). 40 & 100 Gb/s optical communications systems based on blind support vector machine with electrical equalizer. *Journal of Optoelectronics and Advanced Materials, 19*, 146–152.

Niu, D., Wang, Y., & Wu, D. D. (2010). Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications, 37*(3), 2531–2539.

Tasdighi, M., & Kezunovic, M. (2017). Preventing transmission distance relays maloperation under unintended bulk DG tripping using SVM-based approach. *Electric Power Systems Research, 142*, 258–267.

Türkay, B. E., & Demren, D. (2012). Electrical load forecasting using support vector machine. *International Conference on Electrical and Electronics Engineering, 6*, 49–53.

Xia, C., Zhang, M., & Cao, J. S. (2017). A hybrid application of soft computing methods with wavelet svm and neural network to electric power load forecasting. *Journal of Electrical Systems & Information Technology, 6*. doi:10.1016/j.jesit.2017.05.008

Zeng, J., & Qiao, W. (2013). Short-term solar power prediction using a support vector machine. *Renewable Energy, 52*(2), 118–127.

Zhirui, W., & Runing, T. (2013). Air conditioning load prediction and experimental research based on support vector machine algorithm. *Refrigeration Technology, 3*(4), 28–31.

Zhong, Z., Yang, C., Cao, W., & Yan, C. (2017). Short-term photovoltaic power generation forecasting based on multivariable grey theory model with parameter optimization. *Mathematical Problems in Engineering, 7*, 1–9.