Power Output Prediction of Photovoltaic Plant Based on Big Data

BIN Zhang[1, a], CHI Zhang[2,b], WENQING Zhao[3,c *]and GANG Li[4,d]

[1,2]jinbu road, state grid Tianjin electric power company, Tianjin, China

[3,4]Yonghua north street, baoding, hebei province, China

[a]123@zhangbin.com, [b]jbzwq@126.com, [c]waitadamzh@aliyun.com, [d]ligang@126.com

Keywords: big data, smart grid, PLSR, PV plant, output forecast

Abstract. The development of smart grid has a higher requirement for the real-time data acquisition, processing and association analysis of mass data, big data is one of the core technologies. Accurate prediction of the output of PV (Photovoltaic) plant is of great significance to the safety and stability of power system. The output of PV plant is influenced by the light intensity, temperature and humidity, and there is a complex relationship between factors, there will be a large error when this prediction using LSR(Least Square Regression) analysis. On the basis of electric power data platform, this paper proposes a method for predicting the output of PV plant based on PLSR(Partial Least Squares Regression), this method considers parallel advantage of MapReduce, focuses on the independence of concurrent events, can meet the coverage requirements of power big data attribute dimension and reduction. Through the experimental of data simulation prove that this method is feasible in the treatment of big power data attribute reduction.

Introduction

The development of smart grid has a higher requirement for the real-time data acquisition, processing and association analysis of mass data, company management and the development of business model innovation raise new requirements for the value of data assets, big data is one of the core technologies. The output of PV plant has obvious fluctuation and intermittent, large scale PV plant will bring great challenges to the security and stability of power system and economic operation[1].Accurate forecast of output power of PV plant is conducive to the power system dispatch department to make reasonable scheduling arrangements, so as to reduce the adverse effect caused by the fluctuation of PV power generation. The reference [2] established a statistical regression model and quantitatively established the relationship between the output of the PV plant and the study factor. The reference [3] established a regression model with temperature, cloud cover, wind speed and other factors factor. At the same time, PLSR set multiple linear regression, principal component analysis and canonical correlation analysis in reference [4].According to the factors that affect the output of PV plant, with a strong correlation between factor, Applying PLSR to the output prediction of PV plant has a strong theoretical and reliability.

PLSR.PLSR is proposed by the Swedish chemist Wold S[5].The method extracts a number of new variables that have the best ability to explain the problem, and is used for regression modeling.

Calculation method of PLSR. Combined with the prediction of the paper (only the factor of the output of the PV plant is considered), the modeling steps of the partial least squares regression analysis method are introduced in this paper.

1) Data standardization.

$$\begin{cases} E_0 = (\frac{x_{ij} - \bar{x}_j}{s_j})_{p\times m} \\ F_0 = (\frac{y_i - \bar{y}}{s_y})_{p\times 1} \end{cases} \quad (1)$$

© 2015. The authors - Published by Atlantis Press

Where, E_0 indicates the normalized X matrix; F_0 indicates the normalized Y matrix; x_{ij} represents the value of the i-th sample point of the j-th component of X ; $\bar{x}_j$ represents the mean of the j-th component of X ; s_j represents the standard deviation of the j-th component of X ; y_i represents the value of the i-th sample of Y ; $\bar{y}$ represents the mean of y ; s_y represents the Standard deviation of y .

2) Extract the first principal component. Known standard matrix E_0 and matrix F_0 ,The first principal component(u_1) can be extracted from the E_0 ,Because there is only one dependent variable in this paper, no need to carry out F_0 on v_1 regression. The formula is as follows:

$$\begin{cases} u_1 = E_0 W_1 \\ v_1 = F_0 \end{cases} \tag{2}$$

Where W_1 is the first axis of the E_0 , and $\|W_1\| = 1$.

3) PLSR equation. Assume final extracted main components are $u_1, u_2, \mathbf{L}, u_r$, Then regression equation of F_0 about $u_1, u_2, \mathbf{L}, u_r$ is $F_0 = b_1 u_1 + b_2 u_2 + \mathbf{L} + b_r u_r$ (3)

Since $u_i = E_{i-1} W_i (i \in [1, r])$, And can prove $E_{i-1} = E_0 \prod_{i=1}^{i-1} \left(I - W_i a_i^T \right)$ (I is a unit matrix).You can get

$$u_i = E_{i-1} W_i = E_0 \prod_{i=1}^{i-1} (I - W_i a_i^T) W_i = E_0 W_i^* \tag{4}$$

And then $F_0 = E_0 (b_1 W_1^* + b_2 W_2^* + \mathbf{L} + b_r W_r^*) = E_0 B$ (5)

Where, vector B is the coefficient of PLSR equation.

Forecast Case Analysis.

Data set. In this paper, the research of the partial order reduction algorithm use the Java language, and experiment validation is making in the Hadoop platform. Hadoop is an open source implementation of distributed storage and distributed computing, and it uses the Master/Slave framework, which contains a Master node, one or more Slaves nodes, and unified management to the Slaves by the Master. Due to the limited conditions, this paper will use 15 units which memory of computer (workers) cluster is 512 MB as Slaves nodes, the version number of Hadoop is: Hadoop2.2.0. Take a PV plant as the object of study[6].

Factors affecting, the PV plant output includes temperature, irradiance, PV physical characteristics, the geographical location of PV plant and so on. If all the influencing factors are considered when modeling, it is bound to increase the complexity of the model and increase the prediction error. For a particular PV plant, the physical characteristics of the element, plant location and other factors can be reflected by the historical output data of PV plant. The references [7,8] set up a method for predicting PV output based on neural network. Select the output as the input of neural network model the day before the predicted day. In this paper, we study the relationship between the predict day and the day before, using the output of the previous day as an independent variable.

Select the output of the plant and its influence factors in Jan., 2014 as the experimental data. As shown in Table 1, the data from January 2, 2014 to January 15, 2014 are training set, and the data from January 16, 2014 to January 20, 2014 are testing set.

Tab.1 The power output of UQ Center plant and influencing factors

date	x_1 [kWh]	x_2 [$W \cdot m^{-2}$]	x_3 [°C]	x_4 [%]	x_5 [$m \cdot s^{-1}$]	Output of PV Plant y [kWh]
2014.1.2	2740.41	266.62	24.22	58.91	1.4	2352.73
2014.1.3	2352.73	305.15	27.21	57.82	1.45	2669.13
2014.1.4	2669.13	319.41	29.02	53.72	1.46	2696.42
2014.1.5	2696.42	158.54	25.69	62.75	1.44	1365.18
2014.1.6	1365.18	198.96	25.57	68.64	1.19	1772.74
2014.1.7	1772.74	293.95	24.04	57.01	1.94	2546.98
2014.1.8	2546.98	264.68	23.23	53.52	1.77	2171.81
2014.1.9	2171.81	235.73	22.17	57.54	1.43	2030.44
2014.1.10	2030.44	136.18	21.5	65.18	1.31	1436.94
2014.1.11	1436.94	215.76	22.97	58.72	1.13	1858.7
2014.1.12	1858.7	243.46	23.1	57.91	1.31	2325.91
2014.1.13	2325.91	270.16	23	57.63	1.44	2354.26
2014.1.14	2354.26	264.79	23.54	57.21	1.52	2313.91
2014.1.15	2313.91	263.33	23.56	57.7	1.52	2374.9
2014.1.16	2374.9	253.84	23.77	59.78	1.52	2232.94
2014.1.17	2232.94	303.88	24.02	56.61	1.4	2592.66
2014.1.18	2592.66	247.06	23.44	56.94	1.29	2266.31
2014.1.19	2266.31	326.63	24.01	57.7	1.26	2784.48
2014.1.20	2784.48	314.55	25.3	59.99	1.54	2702.59

Using PLSR to establish the output prediction model of PV plant.

1) Extract the first principal component u_1. First, standardize the independent variable matrix X and the dependent variable matrix Y to E_0 and F_0 respectively, and then

$W_1 = [0.1731, 0.7131, 0.2809, -0.5291, 0.3204]^T$

$a_1 = [0.3367, 0.5962, 0.2962, -0.5628, 0.4230]^T$ $\quad b_1 = 0.5618$

$u_1 = E_0 W_1 = 0.1731 x_1^* + 0.1731 x_2^* + 0.2829 x_3^* - 0.5291 x_4^* + 0.3204 x_5^*$ (x_i^* is the i-th column of $E_0, i = 1,2,3,4,5$)

At this point we can obtain regression equations of y^* (y^* is F_0) about u_1:

$$y^* = b_1 u_1 = 0.5681 u_1 = 0.0973 x_1^* + 0.4006 x_2^* + 0.1578 x_3^* - 0.2973 x_4^* + 0.18 x_5^* \tag{6}$$

2) Extracting second component u_2. Substitute $E_1 = E_0 - u_1 a_1^T$ for E_0, substitute $F_1 = F_0 - u_1 b_1$ for F_0, you can extract u_2.

$W_2 = [-0.7153, 0.5109, -0.0672, 0.1473, -0.4484]^T$

$a_2 = [-0.8708, 0.3803, 0.1039, 0.033, -0.4124]^T$ $\quad b_2 = 0.4345$

u_2 =-0.7153 x_1' +0.5109 x_2' -0.0672 x_3' +0.1473 x_4' -0.4484 x_5' (x_i' is the i-th column of $E_0, i = 1,2,3,4,5$)

At this point we can obtain regression equations of y^* about u_1 and u_2:

$$y^* = b_1 u_1 + b_2 u_2 = 0.5681 u_1 + 0.4345 u_2$$
$$= -0.1963 x_1^* + 0.6935 x_2^* + 0.1565 x_3^* - 0.2858 x_4^* + 0.0170 x_5^* \tag{7}$$

3) Establish PLSR. the above calculation shows that, After extracting the components u_1 and u_2, the cross validation was met, At this point the standard regression equation is:

$y^* = -0.1963 x_1^* + 0.6935 x_2^* + 0.1565 x_3^* - 0.2858 x_4^* + 0.0170 x_5^*$

By the inverse process of standardization, the regression equation of the original variable is obtained, $y = 2105.116 - 0.185x_1 + 5.509x_2 + 32.233x_3 - 29.208x_4 + 34.168x_5$ (8)

Using LSR to establish the output prediction model of PV plant. In order to verify the prediction effect of the model based on PLSR, the LSR model is established using the data from 2014.1.2 to 2014.1.15 in Table 1,The regression equation is:

$$y = -253.293 + 0.26x_1 + 9.223x_2 - 24.23x_3 + 14.491x_4 - 118.142x_5 \quad (9)$$

Comparison of prediction results. Table 2 shows comparison of the output prediction results of PLSR and LSR, using the data from January 16, 2014 to January 20, 2014.From table 2, the ability of prediction of PLSR is obviously higher than that of LSR.

Tab.2 Comparison of predicted results of 2 models

date	Actual	PLSR		LSR	
	output[kWh]	Predictive output [kWh]	Relative error [%]	Predictive output [kWh]	Relative error [%]
2014.1.16	2232.94	2136.22	-4.33	2816.1	26.12
2014.1.17	2592.66	2534.7	-2.24	3202.89	23.54
2014.1.18	2266.31	2123.04	-6.32	2804.2	23.73
2014.1.19	2784.48	2616.92	-6.02	3453.97	24.04
2014.1.20	2702.59	2438.77	-9.76	3446.12	27.51

Conclusions

Combined with the technical advantages of large data and requirements of power system application , the value of big electric power data will bring a new opportunity for the construction of smart grid. In this paper, we propose a method to predict the output of PV plant based on PLSR. And the experimental results verify the effectiveness of our method. In future, the next step will be how to eliminate the irrelevant variables and further improve the regression equation while doing big data modeling research, so that it can better deal with the analysis of the big power data.

References

[1] Ming Ding ,Weisheng Wang , Xiuli Wang. Overview of the impact of large scale photovoltaic power generation on power systems [J]. CSEE,2014, 34(1): 1-3.

[2] Dawei Liu ,Wenbo Peng ,Liangliang Bao.Regression analysis of photovoltaic power generation considering air pollution factors [J].Renewable energy,2014, 32(12): 1786-1789.

[3] Xuemei Lv,Zhu Hong ,Jindong Wang.Analysis of the impact of meteorological factors on the amount of photovoltaic power generation [A]. The thirty-first annual meeting of the China Meteorological Society [C], Beijing,2014.

[4] Lifan Mao ,Yuechun Jiang ,Ruihua Long . A long-term power load forecasting based on partial least square regression analysis [J]. power grid technology, 2008, 32(19): 72-77.

[5] Huiwen Wang . Partial least squares regression method and its application [M]. Beijing: National Defense Industry Press, 1999, 178-204.

[6] The University of Queensland. All arrays combined: 2014 yearly power [EB/OL]. Information on:http://www.solar.uq.edu.au/user/reportPower.php?dtra= year&dts =2014-00-00.

[7]Ming Ding ,Lei Wang ,Rui Bi. Short term forecasting model of output power of photovoltaic power generation system based on improved BP neural network [J]. Power system protection and control,2012, 40(11): 94-99.

[8] Jiawei Zhang ,Zijia Zhang .prediction of short term PV system power generation based on PSO-BP neural network[J],Renewable energy, 2012, 30(8): 28-30.