APPEEC2019 1570570868

Two-stage Photovoltaic Power Forecasting based on Extreme Learning Machine and Improved Pointwise Mutual Information

Zhengrong Chen, Yang Hu*
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
hooyoung@ncepu.edu.cn

***Abstract*- Photovoltaic power forecasting plays an important role in the power system planning and contributes to the development of renewable energy. This paper proposed a two-stage forecasting method based on Extreme Learning Machine (ELM) and Improved Pointwise Mutual Information (IPMI), which is responsible for short-term forecasting of the small-scale PV station. The method like a hybrid method requires past measured Numerical Weather Prediction (NWP) data and time series of PV power output as input of the system. During the first stage, PMI algorithm is applied to solve the coupling problem and to determine the weather features that contribute to the PV power output. Locally weighted regression (LOESS) smoothing is used as a nonparametric technique to fit a smooth curve to calculate regression error. Then, historical data will be classified into several groups by affinity propagation (AP) clustering method, combing with principal component analysis (PCA). For the second stage, training models and prediction models based on ELM network will be set up for each group obtained by AP clustering respectively. Numerical results for a small-scale PV plant in Beijing present improvements in forecasting accuracy and computation efficiency when compared to other forecasting methods.**

***Index Terms*—Photovoltaic forecasting, extreme learning machine (ELM), improved pointwise mutual information (IPMI), PCA-AP clustering.**

I. INTRODUCTION

In recent years, solar photovoltaic power is developing rapidly as a clean, sustainable, and environment-friendly energy source. The average annual growth rate of global photovoltaic power generation capacity in 2005-2018 is higher than 30%. Under optimal conditions, the world's solar generation plant capacity could reach up to 1,270.5 GW by the end of 2022 [1]. It's undoubted that solar is the fastest growing power generation source. However, due to the volatility and randomness of photovoltaic power, increasing photovoltaic power generation has a negative impact on the power grid when it is connected to the grid [2]. Accurate and reliable PV power forecasting is one of the key technologies to solve above issues, especially for short-term prediction. At present, it has become an indispensable part of PV power generation grid-connected management, conventional power plant scheduling and energy market decision-making [3].

Various approaches for solar power forecasting are reported in the literature. There are three main methods: physical method, statistical method and machine learning method, and hybrid method [4]. In many literature, NWP data is essential to predict PV power output. From the perspective of physical power generation of photovoltaic panels, solar irradiance is the main factor determining photovoltaic power, but there are many other factors such as temperature, wind speed, cloud, geographic coordinates, etc., and the factors are coupled with each other. Therefore, the specific factors affecting the PV power should be considered to improve the prediction accuracy.

Statistical and machine learning method as data-driven technology are promising ones [5-6]. It consists of differential autoregressive moving average model (ARIMA), artificial neural network, Markov chain, time series, support vector machine, wavelet analysis algorithm, and random forest algorithm. ELM has been successfully used to predict the output power of photovoltaic power plants [7,8]. Hybrid models are the combination of two or more forecasting techniques to improve the accuracy of the forecast. Many studies have showed that integrated forecast methods outperform individual forecast [8-10]. It is founded that ANN is combined with wavelet to develop a new forecasting method [11]. Other authors like [12-14] used other soft computing techniques like GA, fuzzy logic, Quantum based GA, etc. to develop hybrid models. However, most of them compromise the computation efficiency to set up sophisticated network models as to improve the forecasting accuracy, which may degrade the performance of power system in real-time operation.

This paper proposes a two-stage forecasting method that combines ELM and IPMI. During the first stage, IPMI algorithm is applied to extract weather features that contribute

to the PV power output. LOESS smoothing is used to calculate regression error. Then, historical data will be classified into several groups by AP clustering method. Besides, PCA introduces a low latitude feature to represent high latitude to reduce the complexity of clustering and training. For the second stage, training models based on ELM network will be set up for each group obtained by AP clustering respectively. According to the distances between forecast sample and clustering center, each day-based forecast sample will be automatically updated to a specific group. The NWP data and PV power output time series are used as two-stage ELM (TSELM) model input to forecast the PV power. To our best of knowledge, seldom work applies entropy concept into weather factor selection and combined IPMI with ELM for PV power forecasting, which definitely improve the efficiency of factor selecting and enhance the forecasting accuracy. Moreover, several sets of Benchmark datasets are synthesized to verify the validity of the IPMI method. A comprehensive comparative study is tested on real-world datasets of a Beijing PV power plant to confirm the effectiveness of the proposed method.

The paper is organized as follows. Methodology is introduced in Section II; Section □ presents the training and validation results; and finally, Section IV concludes the paper.

II. METHODOLOGY

A. IMPI Algorithm

In 1948, Shannon firstly proposed the theory of information entropy, which applied the concept of entropy in thermodynamics to probability statistics and solved the problem of quantitative measurement of information [15]. Information entropy can be described as:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

where $H(X)$ is the information entropy of X, p_i is the probability distributions, n is the number of datasets. The part related to the two variables are defined as Mutual Information (MI). Information entropy of mutual information (MI) is

$$I(X,Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{2}$$

$$\hat{f}(x) = \frac{1}{n(\sqrt{2\pi}h)^d \sqrt{|\Sigma|}} \exp\left(\frac{-\|x - x_i\|}{2h^2}\right) \tag{3}$$

where $p(x)$, $p(y)$ are the marginal probability density function (PDF) of X and Y, h is the kernel function parameter, $\sum$ is the variance of x.

For multi-input systems, the coupling relationship between the input would confuse the MI value between input and output. Therefore, the Partial Mutual Information (PMI) is used as the basis for selecting input variables. To improve the accuracy of the variance selection, LOESS as nonparametric learning algorithm is proposed to calculate the regression result as the following:

$$\min \quad J = \sum_i \omega_i (y_i - \theta^T x_i)^2 \tag{4}$$

$$m_y(x) = \theta^T x + R \tag{5}$$

where, ω_i is weight; θ is regression matrix; $m_y(x)$ is the regression value; R is the error which is trivial and can be ignored.

Detailed procedure of IPMI: Let the input variable set be C, the output variable be Y, the optimal input variable set be S, and X_s be the candidate variable corresponding to the maximum PMI value.

Step 1: Initialize S, as an empty set.

Step 2: When $X \neq \emptyset$.

Step 3: Calculate $u = Y - m_y(S)$.

Step 4: For each input variable, calculate $I(X_j, u)$. Find the X_j corresponding to the maximum I as X_s.

Step 5: Calculate the AIC value. If the AIC value decreases, move X_s into S and return to step 2; otherwise, terminate the screening.

$$AIC = n \log\left(\frac{1}{n}\sum_{i=1}^{n} u_i^2\right) + 2 * p^2 \tag{6}$$

Step 6: Return the optimal input variable set S.

where n is the number of samples; p is the number of selected variables; u_i is the regression residual of Y calculated from the selected variables. When the AIC reaches the minimum value, the optimal independent variable set is extracted.

Benchmark

This paper synthesizes six sets of Benchmark datasets to verify the validity of the PMI method. Although the datasets are synthetic, they represent a series of different degrees of nonlinear or continuous time series, shown as Table □.

TABLE I. BENCHMARK DATA-GENERATION MODELS

Linear auto-regressive time-series	
AR4	$x_t = 0.6x_{t-1} - 0.4x_{t-4} + \varepsilon_t$
AR9	$x_t = 0.3x_{t-1} - 0.6x_{t-4} - 0.5x_{t-9} + \varepsilon_t$
Non-linear threshold auto-regressive time-series	
TAR1	$x_t = \begin{cases} -0.9x_{t-3} + 0.1\varepsilon_t & \text{if } x_{t-3} \le 0 \\ 0.4x_{t-3} + 0.1\varepsilon_t & \text{otherwise} \end{cases}$
TAR2	$x_t = \begin{cases} -0.5x_{t-6} + 0.5x_{t-10} + 0.1\varepsilon_t & \text{if } x_{t-6} \le 0 \\ 0.8x_{t-10} + 0.1\varepsilon_t & \text{otherwise} \end{cases}$
Non-linear input-output functions	
Firedman	$y = 5(2\sin(\pi x_1 x_2)) + 4(x_3 - 0.5)^2 + 2x_4 + x_5) + \varepsilon$
Mexican Hat	$y = \frac{\sin(\sqrt{x_1^2 + x_2^2})}{\sqrt{x_1^2 + x_2^2}} + \varepsilon$

where, x~N(0,1) is the input variable; ε~N(0,1) is the noise signal. For the first 4 sets of data sets, output variables are x_t, and 15 input variables are x_{t-i}(i=1,2,⋯,15). For the Friedman dataset, the output variable is y, and the input variable is x_1,x_2,x_3,x_4,x_5 and 10 white noise sets; For the Mexican Hat dataset, the output variable is y, and the input variable is x_1,x_2 and 13 white noise sets. The sample sizes of the six

Benchmark data sets are prescribed as 50, 100, and 500; then using IPMI, 30 simulations are performed in each case.

TABLE II. BENCHMARK VERIFICATION RESULTS

	Accuracy rate/%		
Sample size	N=50	N=100	N=500
AR4	50	83.3	90
AR9	43.3	63.3	83.3
TAR1	20	33.3	56.7
TAR2	36.7	70	76.7
Friedman	26.7	33.3	36.7
Mexican	66.7	86.7	86.7

It can be seen from Table II that for linear autoregressive models and nonlinear autoregressive models, as the number of samples increases, the accuracy rate of IPMI screening increases. The candidate variables of the Friedman and Mexican datasets are not related to each other. The screening accuracy rate is basically not affected by the sample. The results shown in the experiment demonstrated that this method can be applied into the selection of weather features.

B. PCA-AP classification

PCA is the most common dimensionality reduction algorithm. When projecting all the data onto the low-dimensional plane, the goal is to minimize the projected mean squared error. The input of the prediction model is generally the weather feature matrix. The data sample is described for one day, the totally three-dimension matrix can reduce a two-dimension matrix by PCA. Then, the AP clustering algorithm is used to classify the weather factor data in days, adopting Euclidean distance as a measure of similarity. With the advantage of AP clustering, there is no need to set the number of classes in advance. A desirable classification result can be obtained after iterations.

For each data point in the sample set, a matrix $S(i,j)$ is used to present the similarity between the data points. It shows the ability of data point x_j becoming the clustering center of data point x_i. The larger the $S(i,j)$ value, the closer distance between the two points. $S(i,j)$ is calculated as:

$$S(i,j) = -\left\| x_i - x_j \right\|^2 \tag{7}$$

Attraction information, $r(i,k)$, represents the degree to which data point x_k fits the cluster center of the data point x_i; membership information, $a(i,k)$, describes the possibility of data point x_i selecting data point x_k as its cluster center. The equation of $r(i,k)$ is shown in formula (8), and the equation of $a(i,k)$ is shown in formula (9)

$$r(i,k) \leftarrow S(i,k) - \max_{k' \neq k} \{a(i,k') + s(i,k')\} \tag{8}$$

$$a(i,k) \leftarrow \min\{0, r(k,k) + \sum_{i' \notin \{i,k\}} \max\{0, r(i',k)\}\} \tag{9}$$

The iterative process computes automatically and terminates until it reaches the maximum iteration times, or the iterative process converges.

C. ELM network

Huang G-B et al. proposed a single-hidden layer feed forward neural networks called ELM [7,16], with structure shown in Figure 1. It's able to produce good generalization performance and learn thousands of times faster than networks trained using backpropagation. This paper applies ELM to approximate the dynamic nonlinearity of photovoltaic power generation.

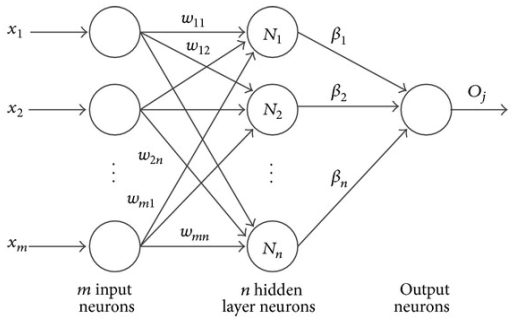

Figure 1. The structure of ELM network

For N data samples $(\mathbf{x}_i,\mathbf{t}_i)$, where $\mathbf{x}_i=[x_{i1},x_{i2,\cdots},x_{in}]^{\mathrm{T}} \in \mathbf{R}^n$ and $\mathbf{t}_i=[t_{i1},t_{i2,\cdots},t_{im}]^{\mathrm{T}} \in \mathbf{R}^m$ are the input and ideal outout, respectively. The activation function $g(x)$ with L hidden neurons are mathematically modeled as follows:

$$\mathbf{o}_i = \sum_{j=1}^{L} \boldsymbol{\beta}_j g(\mathbf{w}_j \cdot \mathbf{x}_i + \mathrm{b}_j), \quad i = 1,2,\ldots,N \tag{10}$$

where $\mathbf{w}_j=[w_{j1},w_{j2,\cdots},w_{jn}]^{\mathrm{T}}$ is the weight vector describing the connection between the jth hidden node and the input nodes; $\boldsymbol{\beta}_j \in \mathbf{R}^m$ is the weight vector describing the connection between the jth hidden node and the output nodes. Using $g(x)$ to approximate those N samples with zero error, the mathematical formula among $\boldsymbol{\beta}_j$, $\mathbf{w}_j$, and b_j satisfies:

$$\sum_{i=1}^{N} \left\| \sum_{j=1}^{L} \boldsymbol{\beta}_j g(\mathbf{w}_j \cdot \mathbf{x}_i + \mathrm{b}_j) - \mathbf{t}_i \right\| = \sum_{i=1}^{N} \left\| \mathbf{o}_i - \mathbf{t}_i \right\| = 0 \tag{11}$$

Equation (11) can be rewritten simply as formula (12):

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \tag{12}$$

where the hidden layer output matrix $\mathbf{H}$ is defined as:

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + \mathrm{b}_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + \mathrm{b}_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + \mathrm{b}_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + \mathrm{b}_L) \end{bmatrix}_{N*L} \tag{13}$$

At the beginning, the input weights $\mathbf{w}_j$ and biases b_j are randomly assigned, then we compute the output weight vector $\boldsymbol{\beta}$ as formula (14) to fulfill the ELM training.

$$\boldsymbol{\beta} = \mathbf{H}^{-}\mathbf{T} \tag{14}$$

where $\mathbf{H}^{-}$ stands for the Moore-Penrose inverse of matrix $\mathbf{H}$ [16].

D. TSELM method

The schematic diagram of the proposed TSELM model is shown in Figure 2. In this proposed forecasting technology, both historical data and weather report information are used in the forecasting process. In the first stage, PV power time series and NWP data as inputs are sent to the IPMI block,

extracting main weather features. For simplicity and computation improvement, 3-D arrays are flattened into 2-D plane to obtain dimensionality reduction by PCA. Then through AP clustering, former 2-D vectors are analyzed and classified into several groups in the units of one day, which means that each arbitrary day can find the attaching group with the similar forecasting model. In addition, the results will be checked whether it is related to some specific relationship (e.g. season or weather type). For the second stage, forecast sample is sent to the model selection block. The model type is updated by calculating the distance between each group center, where the shortest one will be regarded as the training model. Based on former clustering results, historical data remarked by the model type and forecast sample data are sent to the ELM to forecast the day ahead PV power output.

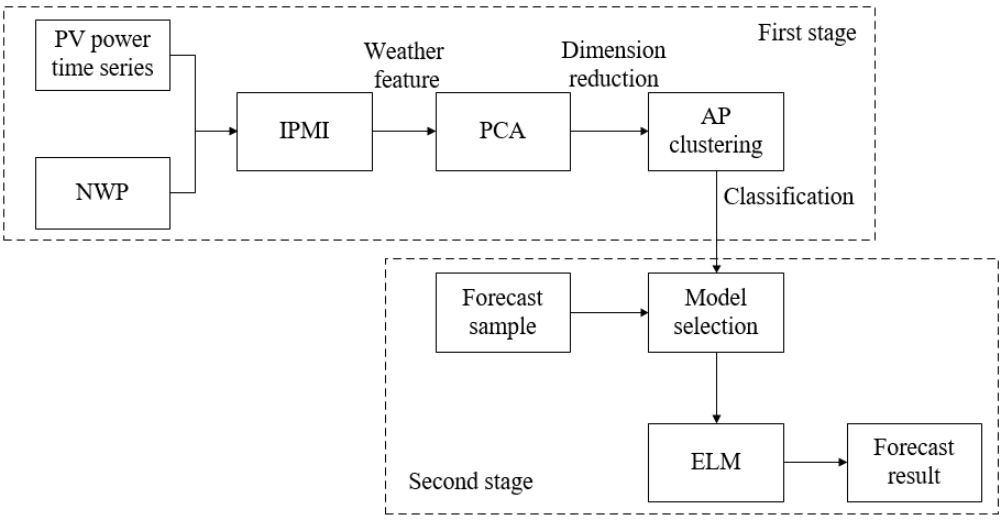

Figure 2. Detailed procedures for the TSELM framework

III. Training and Validation of Results

The NWPs are received from the meteorological services of local weather station, including the daily solar irradiance, the lowest air temperature, the highest air temperature, the daily relative humidity, the daily wind speed, the wind direction, the cloud amounts, the air pressure, and so on. In this paper, the NWPs data is detected from a small-scale PV station in Beijing. The data samples are historical data from 2016/10/21 to 2017/10/22.

A. Data processing

To eliminate the influence of different variable values and dimensional differences on analysis and modeling, the sampled data is normalized as shown in equation (19):

$$z_{\text{norm}} = \frac{z - z_{\min}}{z_{\max} - z_{\min}} \tag{15}$$

where, z_{norm} is the normalized value; z is the original data; $z_{\max}$ is the maximum of data; $z_{\min}$ is the minimum of data.

B. Evaluation metrics

In order to assess the performance of the proposed approach, three evaluation metrics, MAPE, MAE and RMSE are computed. The smaller the error measures, the better the performance of the predictor.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{t_i - T_i}{T_i}\right| \tag{16}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|t_i - T_i\right| \tag{17}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(t_i - T_i\right)^2} \tag{18}$$

where N is the number of samples, T_i is the actual value, t_i is the predictive value.

C. Feature selection and clustering result

Irradiance and temperature, which track the solar state, exhibit significant daily periodic and seasonal variations. Here, a demonstration photovoltaic power station in Beijing is selected as the research object. PV power generation distribution in a whole year is shown in Figure 3.

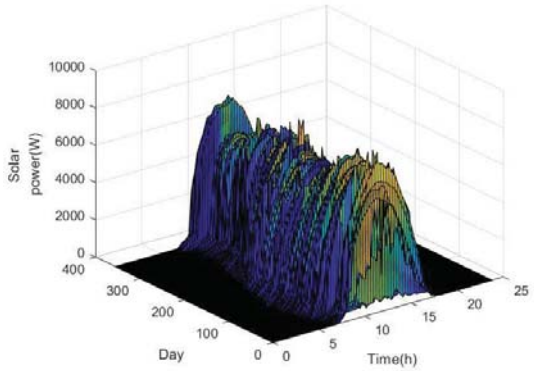

Figure 3. Solar power output curve in three-dimension

From Figure 3, we can see the correlation among different days, and the curve can be utilized for one-day PV power forecasting. Then, IPMI is used to select main weather features. The inputs contain 6 variables: solar irradiance, ambient temperature, relative humidity, wind speed, wind direction and precipitation. Probability distribution and selection of IPMI variables are shown in Figure 4 and Figure 5, respectively. It can be concluded that 4 variables, solar irradiance, ambient temperature, relative humidity and wind speed, are the dominant factors affecting photovoltaic power generation, which are selected as input variables for the next model clustering.

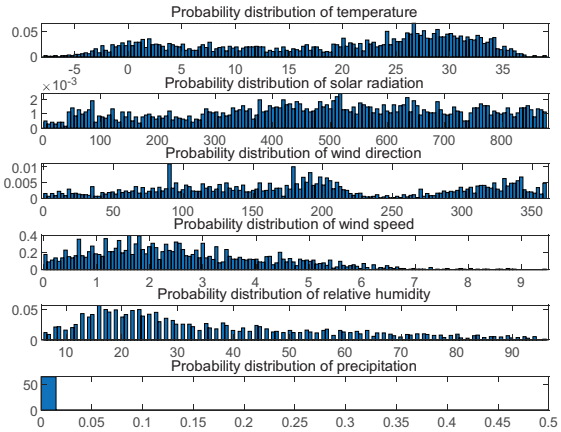

Figure 4. Probability distribution of weather feature

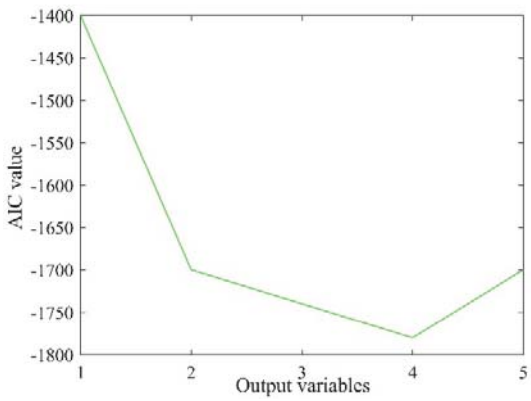

Figure 5. IPMI feature selection output

The result of AP clustering by days is shown in Figure 6. The data available for totally 357 days can be divided into 4 categories, 84 days in group 1, 74 days in group 2, 114 days in group 3, and 85 days in group 4. We can see that the data in each group is not continuous and the two adjacent days may not be in the same group. Then, the prediction model is established for each group, and the historical NWP data is used to perform parameter optimization and determination offline for the next forecasting.

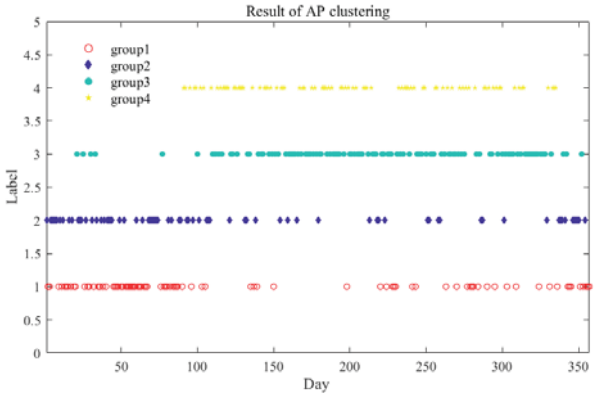

Figure 6. Classification result of AP algorithm

D. *Forecasting result*

In this part, case study was designed to forecast the power output of PV plant for three successive days from 2017/10/23 to 2017/10/25. The hidden nodes of TSELM layer is optimized to be 200. Besides, we compared the result of the proposed TSELM method with that of SVM method and BP neural network to verify the effectiveness of TSELM. According to the time of sunrise and sunset, we selected the data from 7:45am to 4:45pm. The forecasting result is shown in Figure 7. Table III lists the MAPE, RMSE, MAE and training time values of the three compared models, in 72 hours.

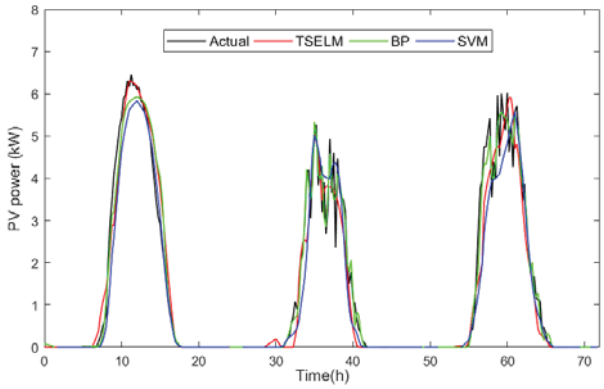

Figure 7. PV power forecast result in 72 hour-ahead

TABLE III. COMPARATIVE ANALYSIS OF TSELM WITH BP AND SVM

	MAPE, %	MAE	RMSE	Training time
TSELM	3.467	0.380	0.0467	00:00:23
BP	4.247	0.253	0.0708	00:03:26
SVM	5.016	0.385	0.0812	00:01:48

From Figure 7 and Table III, it can be observed that the proposed TSELM method outperforms with least MAPE, RMSE and training time among these three. With the advantages of selecting the main weather features and classification for different models, the proposed method shows an increase in forecasting performance and improves the training efficiency by using ELM. The training time in this case is 23 seconds, which is greatly reduced for network building and training, making it a promising option to be used for real-world applications in smart grids. Therefore, it can be

concluded that the proposed TSELM is suitable for short-term PV power forecasting of the small-scale PV power plants.

IV. Conclusion

A novel two-stage PV power solar forecasting method based on ELM and IPMI is proposed in this paper. Firstly, four weather features are extracted by using IPMI. The time series of PV output and NWP in the experiment period are classified into four groups in days by PCA-AP clustering algorithm. Then, the time series of PV output and NWP data for each group are used as the input of ELM network to train the different models. Case study forecasted the power output of PV plant for 3-day ahead from 2017/10/23 to 2017/10/25. Numerical results show that the proposed method has better performance with less error in terms of better prediction accuracy and faster computation efficiency. Due to the highly variable capability, this method could become a part of real-time operation in power system related to numerous issues e.g. energy management, power dispatch, system control and energy market decision-making.

V. Acknowledgment

This work is supported by National Natural Science Foundation of China (U1766204), Fundamental Research Funds for Central Universities (2019MS024), the Beijing Municipal Science and Technology Project (No. Z181100005118005).

References

[1] European Photovoltaic Industry Association. Global market outlook for photovoltaics 2018-2022[R]. EPIA Report, 2018.

[2] R. Varma, M. Salama. Large-scale photovoltaic solar power integration in transmission and distribution networks, *Proceedings of IEEE power and energy society general meeting*, Detroit,2011.

[3] M. G. Villalva, T. D. Siqueira, E. Ruppert, Voltage regulation of photovoltaic arrays: small-signal analysis and control design. *IET Power Electronics*,2010,3(6):869-880.

[4] H. M. Diagne, M. David, P. Lauret, J. Boland and N. Schmutz., Review of solar irradiance forecasting methods and a proposition for smallscale insular grids‖, Renew. Sustain. Energy Rev., vol. 27, pp. 65–76, Nov,2013.

[5] P. Bacher, H. Madsen, H. A. Nielsen, Online short-term solar power forecasting. *Solar Energy*, 2009, 83(10):1772-1783.

[6] W. Xu, "The designing of establishing modeling simulation and fore-casting methods for short-term solar irradiance based on time-scale," Changzhou Inst. Technol., vol. 22, no. 1, pp. 41–44, 2009.

[7] G. B. Huang, H. Zhou, X. Ding, et al. Extreme learning machine for regression and multiclass classification. Systems, Man, and Cybernetics, Part B: *IEEE Transactions on Cybernetics*, 2012, 42(2): 513-529.

[8] I. Majumder, M. K. Behera and N. Nayak, "Solar power forecasting using a hybrid EMD-ELM method," 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Kollam, 2017, pp. 1-6.

[9] L. F. Zarzalejo, L. Ramirez, and J. Polo, Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index‖, Energy, vol. 30, no. 9, pp. 1685 – 1697, 2005.

[10] S. Cao, and J. Cao, Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis, Applied Thermal Engineering, vol.25, no.2-3, pp.161 – 172, 2005.

[11] N. A. Abdullah, S. Koohi-Kamali and N. A. Rahim, "Forecasting of solar radiation in Malaysia using the artificial neural network and wavelet transform," 5th IET International Conference on Clean Energy and Technology, Kuala Lumpur, 2018, pp. 1-8.

[12] W. Jianping, X. Yunlin, Z. Chenghui and X. Xiaobing., Daily solar radiation prediction based on Genetic Algorithm Optimization of Wavelet Neural Network, Int. Conf. on Electrical and Control Engineering (ICECE), Yichang, pp.602-605, Sep. 2011.

[13] S. Hussain and A. Al Alili., Soft Computing approach for solar radiation prediction over Abu Dhabi, UAE: A comparative analysis, IEEE Conf. on Smart Energy Grid Engineering (SEGE), Oshawa, ON, pp.1-6, Aug. 2015.

[14] D. K. Chaturvedi, Forecsting of Solar Power using Quantum GAGNN, Int. J. of Computer Applications, vol. 128, no. 3, pp. 15-19,2015.

[15] C. E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 1948, 27(3): 379-423.

[16] G. B. Huang, Q.Y. Zhu, C.K. Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, 2004:985-990 vol.2.