Stacked-Locally Weighted Ensemble Learning for Wind Speed Prediction

Hong Shi, Man Yu, and Qinghua Hu

Abstract—**It is a key problem to reach a balance between wind power supply and demand in wind power grid. However, the balance is difficult to achieve due to the random and unstable characteristics of wind speed. The accurate wind speed and power forecast model is needed to be developed. In this paper, we design a stacked-locally weighted ensemble method (SLWE) for wind speed prediction based on two facts, wind speeds are typical of the time series; the neighboring wind speeds are more relevant. The previous time prediction is treated as one attribute in the prediction of next time wind speed. The architecture of SLWE not only considers the diversity but also the accuracy. The proposed method is tested on wind speed datasets from Jilin, China. Experimental results show the effectiveness of SLWE.**

Index Terms—**Ensemble learning, local learning, stacked, wind speed prediction.**

I. Introduction

In electricity generation, wind energy has become one of the most rapid growing renewable energy as it is socially beneficial and environmentally friendly. It is a key problem to reach a balance between wind power supply and demand. However, the balance is difficult to obtain due to the random and unstable characteristics of wind speed. A number of effective researches have been proposed. The approaches are roughly divided into physical methods, statistical methods and ensemble methods.

Physical methods, like Numerical Weather Forecast, provide the predictions by considering multiple physical conditions. However, physical conditions are difficult to obtain [1], [2].

For wind speed prediction, statistical methods, such as time series, kernel methods, and artificial neural network (ANN), have been widely used. In 1991, Daniel *et al*. first introduced a linear method, Regressive Moving Average model, to hourly average wind speed prediction [3]. However, wind speeds are typical nonlinear time series. In 2010, Li *et al*. compared three ANN methods (linear element, back propagation and radial basic function) in 1-h-ahead wind speed prediction [4]. Due to robustness of recurrent neural network, Zaccheus *et al*. proposed stacked-recurrent neural network for wind speed prediction [5]. Ramasamy applied temperature, air pressure, solar radiation and altitude as inputs for ANN model to predict wind speeds. The ANN methods have achieved effective results [6]. In 2004, Mohandes *et al*. applied support vector machine (SVM) for wind speed prediction. The experiments demonstrate that the performance of SVM is better than Multilayer Perceptron in wind speed prediction [7]. In 2010, Extreme Learning Machine was introduced into SVM for wind speed prediction (ELM). The model gets fast convergence rate [8]. Hu and Zhang put forward noise-aware SVR models for wind speed prediction in 2014 [9]. The noise-aware SVR model is much more robust. The above nonlinear methods achieve better performance than linear ones.

To improve the robustness and accuracy, ensemble learning is introduced. In 1999, the first ensemble strategy for wind prediction is proposed [10]. The strategies, Basic Ensemble Method (BEG) [11], MLP neural network [12] and mixture density neural network [13], were used for wind speed prediction. Ren *et al.* combined ANN and SVM by empirical mode decomposition method [14]. However, the weighted coefficients obtained from the above algorithms are constant. The same as single models, the idea of constant weight models is to find the average state of the whole historical data.

Affected by atmosphere pressure, there uncertain and unstable characteristics in wind speed. Chinese meteorological department provides 5 typical patterns [15]. Using one global model can't reflect the character of wind speeds. Hu and Su clustered the wind speed into different clusters through generalized principal component analysis and built the wind speed predicting model for each cluster in 2014 [16]. The experiments showed effective results. However, the number of clusters is difficult to obtain, and the clusters are difficult to define. In 2010, Acar *et al*. introduced a dynamic weight fusion method. The method finds some similar wind speed data of each test point and builds the prediction models for each test sample [17]. The experimental results show the effectiveness of the dynamic weight method. However, the number of similar data is not easy to be determined. Both the dynamic weight and cluster-based methods are not flexible for the complex task, and the methods are time-consuming. In 1992, Leon Bottou and Vladimir Vapnik proposed local learning to the complex problem [18].

Considering the dynamic nature of wind speed, we introduce local learning into ensemble learning in our last work. The base models are selected randomly. However, the establishment of base models is not an easy thing in ensemble learning. Due to the different characteristic of the base models, base models are expected to be diverse. Meanwhile,

Manuscript received October 28, 2015; revised January 15, 2016. This work was supported in part by the National Program on Key Basic Research Project (973 Program) under Grant 2012CB215201.

Hong Shi and Man Yu are with the School of Computer Science and Technology, Tianjin University, Tianjin, China (e-mail: 120888349@qq.com, alxyu@tju.edu.cn).

Qinghua Hu is with Harbin Institute of Technology, Harbin, China (e-mail: qinghuahu@tju.edu.cn).

doi: 10.18178/ijmlc.2016.6.1.571

in order to guarantee a low generation error of ensemble learning, the base models should be fairly accurate. It is difficult to balance diversity and accuracy for base models [19]. In this paper, we propose stacked-locally weighted ensemble method (SLWE), which contains two or more ensemble layers. In the first layer, eight different models are introduced as base modes which are as diverse as possible. In the other layers, the establishment of base models is interesting. Wind speed samples which are neighborhood have a strong correlation. So, when we predict the next m*th* moment wind speed, it is effective to use wind speed prediction of (m-1)*th* moment. Then, the new base models are established on the new training data. We treat the new base models and the previous layer final predictor as the base models of next layer base models. The accuracy of new base models is more accurate than previous layer. The essence of the proposed method is that it not only ensures diversity but also accuracy. The experimental assessment of SLWE is carried out on the real dataset, Jilin wind speed farm. The results show that five base models are enough. SLWE method outperforms other methods, such as single methods and constant weighted ensemble methods.

The remainder of the paper is organized as follows: Section II analyzes the data of wind speed. Section III describes the proposed stacked-locally weighted ensemble method. Section IV presents the experiments. Finally, the conclusion of this work was drawn.

II. Analysis of Wind Speed

In this work, Jilin wind farm which stores the wind speed from 2011 to 2013 is selected. The average wind speed of 10 minutes is recorded. Fig. 1 shows partial wind speed in 2011.

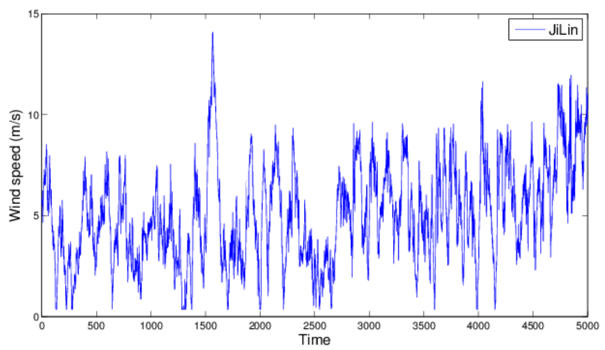

Fig. 1. Wind speed from Jilin wind farm.

From Fig. 1, the most striking characteristic of wind speed is unstable. The wind speed varies from time to time [20]. Wind speed model is complex and nonlinear. It is unsuitable to build wind speed model by global learning.

Chinese meteorological department provides 5 typical wind speed patterns which are created by k-means clustering, shown in Fig. 2 [21]. The wind speed satisfying the same function is clustered into one cluster. The Jilin wind speeds are roughly divided into 5 clusters. It is wise to build wind speed model by local learning.

It is meaningful to compute the correlation in wind speeds. Mutual information is a classic method to calculate the wind speeds. Suppose X and Y are discrete variables with different time. The mutual information is computed as

$$\begin{aligned} I(X,Y) &= H(X)+H(Y)-H(X,Y), \\ H(X) &= -\sum_{i=1}^{m} p_i \ln p_i, \\ H(X,Y) &= -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \ln p_{ij}, \end{aligned} \tag{1}$$

where p_i is the probability. We calculate correlation of $\{x_{t-k+1},\ldots,x_t\}$ and x_{t+1} as k changes to measure the correlation in wind speed time, where x_t is the wind speed at time t. The result is given in Fig. 3. There is a strong correlation between the neighboring wind speeds.

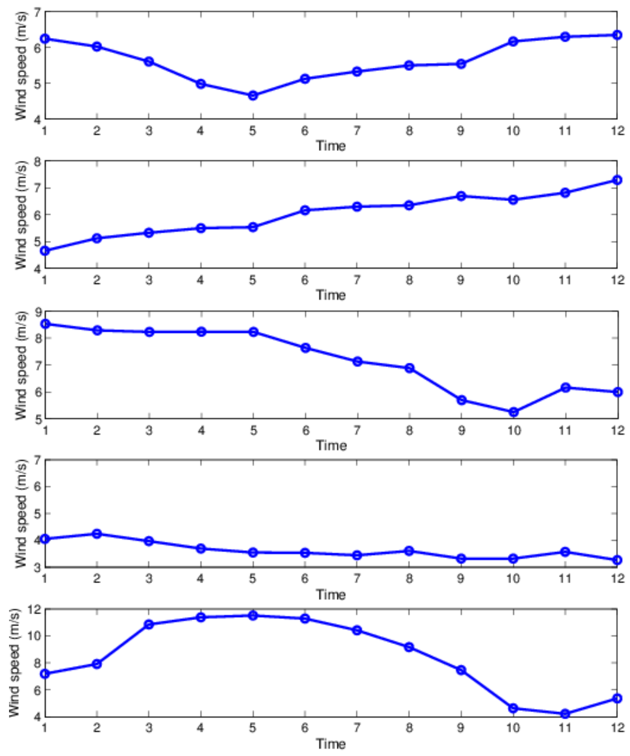

Fig. 2. 5 typical wind speed patterns.

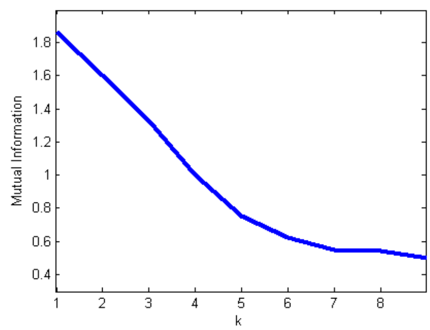

Fig. 3. Mutual information with the change of *k*.

III. Model Framework

In this section, the framework of the proposed method, SLWE, is introduced. We first try to describe how to select the input variables for wind speed prediction. Then the architecture of SLWE is described. We generate the input space as $X_t = (x_{t-11}, x_{t-10}, ..., x_t)$, where x_t is the wind speed at time t. If we predict next m*th* moment wind speed, $y_t = x_{t+m}$. The training set can be drawn as $D = (\langle X_1, y_1 \rangle, ..., \langle X_t, y_t \rangle, ..., \langle X_n, y_n \rangle)$. n is the number of the training samples. For convenience, we take the two-layer locally weighted method as an example.

A. Architecture of Stacked-Locally Weighted Ensemble Learning

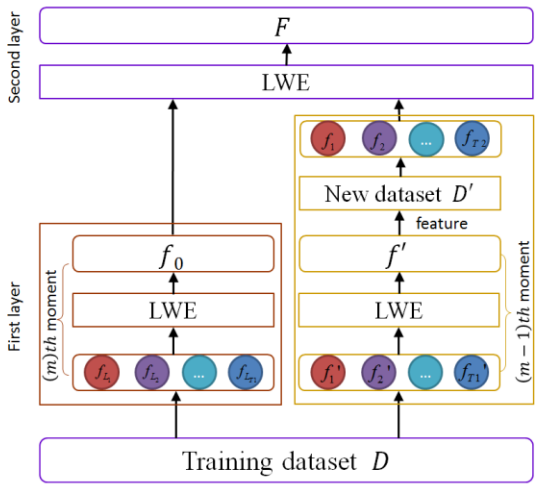

Fig. 4. Architecture of stacked-locally weighted ensemble model.

Fig. 4 illustrates the framework of the stacked-locally weighted ensemble method. The model solves two important problems: How to integrate base models and how to generate base models. For the former one, we have discussed in our former work (LWE). The soft-max function is introduced to the fusion strategy. The fusion strategy conforms to the characteristics of wind speed. The discriminative function of LWE is computed as

$$F(X) = \sum_{i=1}^{M} w_i f_i(X)$$
$$w_i = \frac{\exp(v_i X + v_{i0})}{\sum_{j=1}^{M} \exp(v_j X + v_{j0})}, \quad (2)$$

where $f_i(X)$ is the i_{th} base model and M is the number of base models. The parameters $\langle v_i, v_{i0} \rangle, i = 1, 2, ..., M$ are obtained by

$$\min \frac{\lambda}{2} \|w\|_F + (F(X) - y)^2 \quad (3)$$
$$s.t. \quad w \ge 0$$

For the second one, the base models should be diverse and accurate in ensemble learning. In the following subsections, how to generate base models is described in detail.

B. Establishment of Base Models

In two-layer locally weighted ensemble method, we establish the base models twice. In first layer, the base models are diverse as far as possible. Eight single models (Least square regression (LR), mat_primal, ELM, Feed-forward neural network (FNN), Elman neural network (ENN), Recurrent neural network (LRN), Support Vector Machine Regression (SVR), k-Nearest Neighbor algorithm (KNN)) are selected as one-layer base models, T_1. After intergation, the first predictor f_0 is obtained.

In second layer, previous time prediction is added to ensure the accuracy of base models. Wind speed is typical time series. There is a strong correlation between wind speeds. The previous time wind speed has an impact on current time data which is discussed in Section II. When previous time prediction is obtained, a new training set is generated by adding the prediction to the training data. Then the new training data is applied to model different regression models (ELM, FNN, ENN, LRN), T_2. The final predictors, F, is obtained by combing T_2 and f_0.

C. Algorithm of Stacked-Locally Weighted Ensemble Learning

In our former paper, LWE, provides a dynamic ensemble method. This method depends on the characteristic of wind speed. For accurate wind speed prediction, we formulate the algorithm of SLWE when we predict the next m_{th} moment wind speed at time t. The important value of the proposed architecture considers both diversity and accuracy of base models. Here the algorithm of SLWE is proposed.

Algorithm Stacked-locally Weighted Ensemble algorithm

1: Input: Training data D
2: Output: Final predictor F
Begin:
3: First layer: Train a set of different models, $\{f_{L1}, f_{L2}, ..., f_{L_{T1}}\}$ and get first layer predictor, f_0, through LWE;
4: Other layer: Train models, $\{f_1', f_2', ..., f_{T_1}'\}$, at (m-1)*th* moment wind speed and get new predictor f' through LWE;
5: Get new training data D' of m*th* moment speed by adding new feature of f';
6: Train different models $\{f_1, f_2, ..., f_{T_2}\}$ on D';
7: Get final predictor F by combing $\{f_1, f_2, ..., f_{T_2}\}$ and f_0 with LWE.
End

IV. Experiment and Analysis

In this section, we present the performance of SLWE on Jilin wind farm. We extract 103680 wind speed samples (two years) as training data and 47520 wind speed samples (eleven months) as test data. At time t, we generate the input space as $X_t = (x_{t-11}, x_{t-10}, ..., x_t)$, where x_t is the wind speed at time t. We set the number of attributes as 12. If we predict next (m)*th* moment wind speed, the output is $y_t = x_{t+m}$. In this paper, the next one-hour wind speed is tested.

We compare SLWE with individual regression method (LR, mat_primal, ELM, FNN, ENN, LRN) and global ensemble method (BEG fusion strategy, linear fusion strategy and LWE) [14]. The criteria, mean absolute error (MAE) and root mean square error (RMSE), are used in wind speed prediction [22]

$$MAE = \frac{1}{n_t}\sum_{i=1}^{n_t}\left|F\left(X_i\right)-y_i\right|,\qquad RMSE = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}\left(F\left(X_i\right)-y_i\right)^2} \tag{4}$$

where y_i is the real value of wind speed , $F\left(X_i\right)$ is the i_{th} final predicted value and n_t is the number of the test samples.

A. Analysis of Base Models

In order to gain further insight into local behavior of wind speed prediction, we display the MAEs of first ten days in January, shown in Fig. 5. The base models are diverse. In the same day, MAE is different with the different models. In different days, the performances of different models are diverse. For example, the performances of mat-primal and ELM are relatively good in the first day. But, in the second day, the performance of mat-primal is the worst among the six models. Neither the constant weight ensemble method nor global single model reflects the characteristics of wind speed.

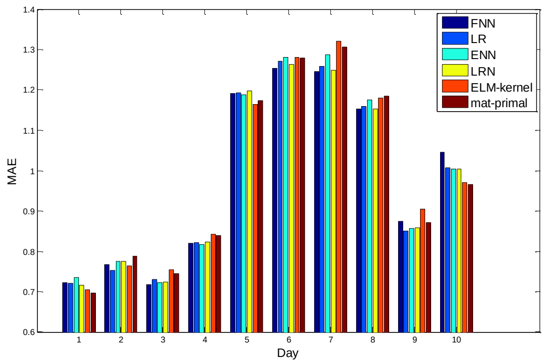

Fig. 5. One-hour MAEs of individual models in first ten days in January.

B. Selection of Base Models

Fig. 6 shows the variation of MAE, RMSE and training times with the number of base models of the first layer. MAE and RMSE seem to be stable when we provide 4 or 5 models. With 5 base models, MAE has fallen by 0.096 and RMSE has fallen by 0.091. However, the consuming times grows linearly with the number of base models. From Fig. 6, five base models are enough. In the next layer, we only use ELM, FNN, ENN, LRN and previous time prediction as the base models.

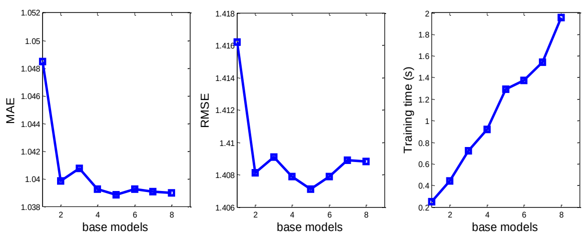

Fig. 6. MAE, RMSE and training time on one-hour wind speed prediction.

C. Prediction of Wind Speed

It is curious why SLWE is more effective than first layer prediction. In Table I, we see the performance of first layer base models and the last layer base models. Obviously, corresponding MAE and RMSE are lower in the last layer. To illustrate the behavior of SLWE, we compare with single model and the fusion methods with BEG, linear fusion method (LR) and our former work (LWE) in Table II. Our proposed method, SLWE, gets better performance than LRN. MAE is reduced by 0.021.The fusion methods is a litter better than single. Compared with LWE, MAE is reduced by 0.011 and RMSE is reduced by 0.02. This evaluates the effectiveness of SLWE.

TABLE I: First-Layer (F) Prediction and Our Prediction (O) of Base Models

	MAE				RMSE			
	FNN	ENN	LRN	ELM	FNN	ENN	LRN	ELM
F	1.049	1.053	1.048	1.055	1.420	1.419	1.416	1.435
O	1.045	1.049	1.045	1.049	1.413	1.415	1.416	1.435

TABLE II: Comparison of Ours, Single Models and Ensmble Methods

	Single Models		Ensemble Models			Ours
	LR	LRN	BEG	LR-fusion	F-layer	
MAE	1.049	1.048	1.042	1.042	1.038	1.027
RMSE	1.417	1.416	1.416	1.413	1.412	1.392

In Fig. 7, the predictions SLWE and BEG ensemble strategy are shown (upper row). The left side is the medium wind speed prediction (>5, <40 m/s) and the right side is the slow wind speed prediction (<5 m/s). We see that the predicting curve of SLWE is closer to real wind speed than that of BEG. The bottle row shows the errors of BEG and SLWE. The error of SLWE is closer to zero.

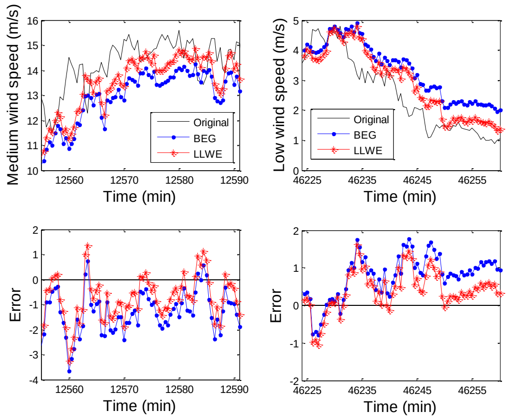

Fig. 7. One-hour wind speed prediction and Errors of BEG and SLWE.

V. Conclusion

In this paper, a novel accurate wind speed prediction method is presented, SLWE. Regression methods, such as FNN, LR, ENN, LRN, ELM, mat_primal, SVM and KNN, are selected to define base models. A number of experiments on Jilin wind farm confirm the effectiveness of SLWE.

1) Wind speed varies from time to time. Wind speed sequence underlying has different patterns. It is not wise to use one global union model for wind speed prediction. Try to model wind speed from local method.
2) The important value of the proposed architecture considers both diversity and accuracy of base models. We

introduce stacked-locally weighted ensemble method to ensure the diversity and accurate.

ACKNOWLEDGE

This work was supported by National Program on Key Basic Research Project (973 Program) under Grant 2012CB215201, National Natural Science Foundation of China under Grant 61222210 and New Century Excellent Talents in niversity(NCET) under Grant NCET-12-0399.

REFERENCES

[1] M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou, "Short term forecasting of wind speed and related electrical power," *Solar Energy*, vol. 63, no. 1, pp. 61-68, 1998.

[2] A. Kusiak, H. Zheng, and Z. Song, "Wind farm power prediction: A data-mining approach," *Wind Energy*, vol. 12, no. 3, pp. 275-293, 2009.

[3] A. R. Daniel and A. Chen, "Stochastic simulation and forecasting of hourly average wind speed sequences in Jamaica," *Solar Energy*, vol. 46, no. 1, pp. 1–11, 1991.

[4] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313-2320, 2010.

[5] Z. O. Olaofe, "A 5-day wind speed & power forecasts using a layer recurrent neural network," *Sustainable Energy Technologies & Assessments*, vol. 6, pp. 1-24, 2014.

[6] P. Ramasamy, S. S. Chandel, and A. K. Yadav, "Wind speed prediction in the mountainous region of India using an artificial neural network model," *Renewable Energy*, vol. 80, pp. 338-347, 2015.

[7] M. A. Mohandes, T. O. Halawani, and S. Rehman, "Support vector machines for wind speed prediction," *Renewable Energy* , vol. 29, no. 6, pp. 939-947, 2004.

[8] B. Frénay and M. Verleysen, "Using SVMs with randomised feature spaces: an extreme learning approach," in *Proc. 18th European Symposium on Artificial Neural Networks Computational Intelligence and Machine Learning*, 2010.

[9] Q. H. Hu, S. G. Zhang *et al*., "Noise model based ν-support vector regression with its application to short-term wind speed forecasting," *Neural Networks*, vol. 57, pp. 1-11, 2014.

[10] T. N. Krishnamurti, "Improved weather and seasonal climate forecasts from multi-model super ensemble," *Science*, vol. 285, no. 5433, pp. 1548–1550, 1999.

[11] L. Bruzzone and F. Melgani, "Robust multiple estimator systems for the analysis of biophysical Parameters from remotely sensed data," *IEEE Trans. on Geoscience Remote Sensing*, vol 43, no. 1, pp. 159–174, 2005.

[12] H. Bouzgou and N. Benoudjit, "Multiple architecture system for wind speed prediction," *Apply Energy*, vol. 88, pp. 2461-2471, 2011.

[13] Z. Men, E. Yee, and F. S. Lien, "Short-term wind speed and power forecasting using an ensemble of mixture density neural networks," *Renewable Energy*, vol. 87, pp. 203-211, 2016.

[14] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 236-244, 2015.

[15] P. Y. Su, "study of patterns-based wind speed prediction," HIT University, 2013.

[16] Q. H. Hu, P. Y. Su, D. R. Yu, and J. F. Liu, "Pattern-based wind speed prediction based on generalized principal component analysis," *IEEE Transactions on Sustaininable Energy*, vol. 5, no. 3, pp. 866-874, 2014.

[17] E. Acar, "Various approaches for constructing an ensemble of metamodels using local measures," *Structural and Multidisciplinary Optimization*, vol. 42, no. 6, pp. 879-896, 2010.

[18] L. Bottou and V. Vapnik, "Local learning algorithms," *Neural Computation*, vol. 4, no. 6, pp. 888-900, 1992.

[19] J. Mendes-Moreira, C. Soares, and A. M. Jorge, "Ensemble approaches for regression: A survey," *Acm Computing Surveys*, vol. 45, no. 1, p. 7, 2012.

[20] D. Jena and S. Rajendran, "A review of estimation of effective wind speed based control of wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 1046-1062, 2015.

[21] M. Beniston and P. Jungo, "Shifts in the distributions of pressure, temperature and moisture and changes in the typical weather patterns in the Alpine region in response to the behavior of the North Atlantic Oscillation," *Theoretical and Applied Climatology*, vol. 71, no. 1, pp. 29-42, 2002.

[22] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983-991, 2008.

Hong Shi received the PhD degree from Harbin Institute of Technology, Harbin, China in 2004. She is currently an associate professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. His research interests include data mining, computer vision and pattern.

Man Yu is currently a master student with the School of Computer Science and Technology in Tianjin University. Her research interests is ensemble learning and data mining.