Wind Speed Patterns Mining Based on Multiple Views

To cite this article: Zongxia Xie and Jimin Jin 2018 IOP Conf. Ser.: Earth Environ. Sci. 168 012032

View the article online for updates and enhancements.

Related content
- DO VOIDS CLUSTER?
  S. Haque-Copilah and D. Basu
- Using Sentinel-1 SAR satellites to map wind speed variation across offshore wind farm clusters
  S F James
- The footprint of atmospheric turbulence in power grid frequency measurements
  H. Haehne, J. Schottler, M. Waechter et al.
Wind Speed Patterns Mining Based on Multiple Views

Zongxia Xie, Jimin Jin

School of Computer Software, Tianjin University, Tianjin, China
e-mail: caddiexie@hotmail.com, jjmtju@163.com

Abstract. There are different patterns for wind speed variation in different seasons. The patterns are not only related to the statistical characteristics, but also to the trends of wind speed fluctuation and variation. However, the classical clustering can’t take the trend information as an independent overall feature in the algorithm. In this paper, multi-view clustering is introduced to extract patterns of wind speed variation from both statistical information and the trends. For about 1.5-year historical data, the trends of wind speed variation and fluctuation, and other five statistical characteristics are selected as features to characterize patterns of wind speed variation. Then four patterns are obtained with multiple clustering. It is found that mean wind speed and fluctuation level are very important and the behavior of wind speed is significantly different in the past states in different clusters. Forecasting model performs worse for the cluster with large mean wind speed and fluctuation range.

1. Introduction

The energy crisis and environmental crisis caused by excessive consumption of fossil fuels become more and more serious. As renewable energy, wind power has been developed rapidly [1]. However, the inaccuracy of prediction for wind speed and wind power restricts the stability of large-scale wind power generation. Two main methods widely used for wind power prediction are numerical weather prediction (NWP) and data driven models. Based on the meteorological formula, NWP can usually obtain stable weather forecasting [2]. However, the calculation cost of NWP is expensive, and the errors in the short-term and the ultra-short term wind speed prediction are large. While data driven models are widely used in short-term and ultra-short term wind speed prediction [3]. The main algorithms are artificial neural networks, SVR [4,5], autoregressive moving average model [6], kalman filtering method [7], recurrent neural network [8], and so on.

As known, wind speed has strong randomness, and wind speed patterns are different in different seasons or days [9]. For these complicated characteristics of wind speed, research work is mainly carried out from two aspects. Firstly, hybrid models are used for wind speed and power prediction. In 2006, Sanchez proposed the short-term prediction of wind energy by assigning different weighting coefficients to several time series models for building hybrid model [10]. Since then, many hybrid models for wind speed prediction have been proposed [11-14]. DL et al decompose wind speed into different components by wavelet technique [12]. Fei et al proposed the hybrid model of empirical mode decomposition and multiple-kernel relevance vector regression algorithm (EMD-MkRVR) to decompose the data into different components by modal decomposition[13]. Ouyang et al improved the accuracy of wind speed prediction by using different kernel functions for different samples with SVR [14]. Secondly, clustering is used to extract the wind speed patterns in advance, and then different models for different patterns are constructed, respectively. The wind speed patterns were
proposed in 2009 [15]. Kusiak and Li obtained different wind speed patterns by using K-means algorithm to cluster the wind speed in different scenes [16]. Hu et al pointed out that K-means which is good at finding the circular cluster is not conducive to explore the patterns of wind speed variation, and proposed generalized principal component analysis for automatically dividing wind speed into different clusters and obtaining wind speed patterns [17]. Lee and Piao proposed the projection clustering to get the patterns of wind speed variation [18]. In 2016, Wani et al used cluster based approach for mining patterns to predict wind speed [19]. However, the existing clustering methods are mainly based on the statistical characteristic of wind speed, such as the mean, variance and so on, or the simple mixture of the trend information and statistical characteristics of wind speed. It is difficult to regard trend information or statistical information as an independent overall feature in the process of clustering. Based on this, we introduce a method based on multiple views for patterns extraction. The method can take the trend information as an independent overall feature in the process of clustering with statistical characteristics at the same time, and make full use of the trend information of wind speed and statistical information to mine wind speed patterns. The method is applied to the real wind speed data, and the latent patterns of wind speed variation are extracted. What’s more, decision tree is used on statistical features to get the rules of different clusters. The boxplots of trend features of different clusters are also analyzed.

2. Analysis on Variation of Wind Speed
The wind speed data is from a wind farm in Ningxia Province located in Northwest China. About 1.5-year historical data is used in this work. The observations of wind speed in the wind farm are in 10-min interval. Figure 1 shows all the wind speed. Figure 2 shows the corresponding first order difference sequence.

As shown in Figure 1, the wind speed is on the rise in some time period. After it climbs to a peak, it slowly or fast drops to a small value. The fluctuation of wind speed is obviously quite large. For certain period of time, the wind speed fluctuates in a certain small range with relatively steady wind speed values. And the mean wind speed is also different at different time periods.

As shown in Figure 2, the wind speed occasionally fluctuates very large, but the wind speed fluctuations are relatively small in most time. The fluctuation range of wind speed is different at different times. The variation of wind speed is very complicated.

For different time periods, the peak value of wind speed, mean wind speed and fluctuation level are also obviously different. These statistical features characterize the overall variation of wind speed. Therefore, it is necessary to consider them as the features to extract the wind speed patterns. The statistical information reflects the general information of wind speed variation. However, the trend
information reflects the detailed information of wind speed variation. In order to incorporate trend information and statistical information for clustering at the same time, we fuse the information from multiple views instead of simply connecting the two types of information.

3. Methods

3.1. Defining the Features for Clustering

According to Part II, both the statistical information and trends of wind speed should be used for clustering. In this paper, two trends of wind speed are collected. One is the trend of wind speed variation, which is composed by the current wind speed \( v_t \) and the previous \( m-1 \) wind speeds \([v_{t-1}, v_{t-2}, \ldots, v_{t-(m-1)}]\). The other is the trend of wind speed fluctuation, which includes \( m-1 \) differences of wind speeds between the two nearest times. At the same time, another five statistical features are computed. They are maximum, minimum and mean wind speed, absolute fluctuation level and fluctuation range of the current and previous \( m-1 \) times. Both the trends and statistical features are listed in Table I.

| Features                          | Definition                                                                 |
|-----------------------------------|---------------------------------------------------------------------------|
| Trend of wind speed variation     | \([v_t, v_{t-1}, v_{t-2}, \ldots, v_{t-(m-1)}]\)                          |
| Trend of wind speed fluctuation   | \([\Delta v_t, \Delta v_{t+1}, \ldots, \Delta v_{t-(m-1)}]\)              |
| Maximum wind speed                | \(\max_{i=0,1,\ldots,m-1} \{v_{i+i}\}\)                                   |
| Minimum wind speed                | \(\min_{i=0,1,\ldots,m-1} \{v_{i+i}\}\)                                  |
| Mean wind speed                   | \(\text{mean}_{i=0,1,\ldots,m-1} \{v_{i+i}\}\)                          |
| Absolute fluctuation level        | \(\sum_{i=1}^{m-1} \text{abs}(\Delta v_i)\)                              |
| Fluctuation range                 | \([v_{\text{max}} - v_{\text{min}}]\)                                   |

In order to use all the features in Table I for clustering, multi-view clustering is introduced. This method can take each trend as a view and all the statistical features as a view. From Table I, we know that there are 3 views for wind speed clustering.

3.2. Multi-view Clustering Algorithm

In this part, the multi-view clustering is introduced to obtain clusters from multiple views. This method is proposed by shao et al [21], which can carry on clustering from multiple views fastly. In general, given a set of incomplete multi-view data \(X^{(\nu)} \in R_{n \times d}, \nu = 1, 2, \ldots, n_o\), our aim is to find the latent feature matrices for each of the views and a common consensus, which represents the integrated information of all the views. The objective function can be written as:

\[
\min_{(x^{(\nu)})_{1 \leq \nu \leq n_o}, (y^{(\nu)})_{1 \leq \nu \leq n_o}} \frac{1}{n_o} \sum_{\nu=1}^{n_o} \|X^{(\nu)} - \hat{U}^{(\nu)}\|_F^2 + \sum_{\nu=1}^{n_o} \alpha_{\nu} \| \hat{Z}^{(\nu)} - Z \|_F^2
\]
\[ s.t. \quad Z^* \geq 0, U^{(v)} \geq 0, Z^{(v)} \geq 0, v = 1, 2, \ldots, n_v \]

where \( U^{(v)} \in R_{n_v \times K} \) is the basis matrix, \( Z^{(v)} \in R_{n_v \times K} \) is latent feature matrix for the \( v \)-th view, \( Z^* \in R_{n \times K} \) is the consensus latent feature matrix, \( K \) is the number of clusters, and \( \alpha_v \) is the trade-off parameter in the objective function.

In order to eliminate the influence of the incomplete data, the instances that appear in the view are given weight 1, and the missing instances in the view are given lower weight. The objective function can be written as:

\[ 1 = \sum_{v=1}^{n_v} \left\| X^{(v)} - U^{(v)}Z^{(v)} \right\|^2 + \sum_{v=1}^{n_v} \alpha_v \left\| Z^{(v)} - Z^* \right\|^2. \quad (2) \]

To enforce the sparsity of the latent feature matrix, \( L_1 \) norm is adopted. The new objective function can be written as:

\[ 1 = \sum_{v=1}^{n_v} \left\| X^{(v)} - U^{(v)}Z^{(v)} \right\|^2 + \sum_{v=1}^{n_v} \alpha_v \left\| Z^{(v)} - Z^* \right\|^2 + \sum_{v=1}^{n_v} \beta_v \left\| Z^{(v)} \right\|_1. \quad (3) \]

where \( \left\| \cdot \right\|_1 \) is \( L_1 \) norm and \( \beta_v \) is the trade-off parameter between the sparsity and accuracy of reconstruction for the \( v \)-th view. After optimization, we can get the best \( Z^* \in R_{n \times K} \), which is the consensus latent feature matrix. Therefore, K-means algorithm is performed on \( Z^* \) to get the latent clusters from multiple views.

As to wind speed clustering, the number of views is 3. All the data of 3 views are used in the objective function (3) to get the optimal \( Z^* \), and then K-means is applied to \( Z^* \) to obtain the final clusters.

4. Experiment

The dataset from a wind farm in Ningxia Province is used in the experiment. To predict the wind speed in the following one-hour, the current and previous \( m \) past states of wind speed are collected as the inputs of wind speed prediction model. \( m \) is set to 23 in this work according to the mutual information [14]. Then the wind speed time series can be transformed into matrix data, each row can be regarded as a sample. The total number of samples is 83483. Among all samples, two-third of the data are used as the training set, and the left one-third as the test set.

To obtain the patterns of wind speed variation, multiple features from three views as shown in Table I are collected at first. Then multiple-view clustering of wind speed samples is conducted on the three views.

4.1. Determining the Number of Clusters

The number of clusters \( k \) is needed to determine in advance for multiple-view clustering. In order to get the suitable number of clusters, \( k \) is set to 2, 3, 4, 5, respectively. Then, SVR is used to construct models on each cluster. The average errors on training data of all clusters with different \( k \) are presented in Table II. The mean absolute error (MAE) and the mean square error (MSE) are computed as:

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_{r,i} - y_{p,i} \right| \quad (4) \]

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left( y_{r,i} - y_{p,i} \right)^2 \quad (5) \]

where \( y_{r,i} \) is the real value, \( y_{p,i} \) is the predicted value, and \( n \) is the number of test samples.
Table 2: Error With Different k Values

| Error | K=2   | K=3   | K=4   | K=5   |
|-------|-------|-------|-------|-------|
| MAE   | 1.0018| 1.0014| 1.0002| 1.0019|
| MSE   | 2.0971| 2.0913| 2.0879| 2.0936|

As shown in Table II, the errors become smaller with the increase of the number of clusters $K$. However, the errors reach the smallest value when $K$ is 4 and then become bigger when $K$ is 5. Therefore, 4 is chosen as the suitable number of clusters.

4.2. Obtaining patterns based on multiple views

Multiple-view clustering is performed on the training data with the number of clusters $K = 4$. The number of samples in each cluster is listed in Table III. Cluster 2 has the most samples, cluster 3 has less samples than cluster 2, while cluster 1 and cluster 4 have the least samples. The sample number of cluster 1 and cluster 4 is only about one-ninth of cluster 2.

In order to get the rules of wind speed patterns, the decision tree is constructed based on the statistical features of these four different clusters. The tree is shown in Figure 3. In addition, some trends of wind speed variation and fluctuation of each cluster are also shown in the leaf nodes by a part of samples.

$\delta$ in Figure 3 represents the mean wind speed in the 24 past state of wind speed, and $\delta$ represents the fluctuation level of wind speed in the past state. And the tree can get 82% accuracy on training data.

We can know that the mean speed and fluctuation level are partitioned into four intervals. For cluster 2, the mean speed is in [0 5.63] and fluctuation level is in [0 10.35]. The mean speeds of cluster 3 are in [0 5.63] and the fluctuation levels are larger than 10.35. The second rule for cluster 3 is that the mean speed is between 5.63 and 10.88, and the fluctuation level is smaller than 13.21. As for cluster 1, the mean speed is also between 5.63 and 10.88, but the fluctuation level is larger than 13.21. The rule for cluster 4 is just that the mean wind speed is larger than 10.88. What’s more, the trends of different clusters are also different.

To better demonstrate the trend rules among the different patterns, the boxplot of samples of each cluster in each 10-min intervals are plotted in Figure 4 and Figure 5. Some outliers are removed to show the main trend. Figure 4 and Figure 5 show that the trends of wind speed variation and fluctuation in different cluster are different. For cluster 1 which has median speed and large fluctuation level, the trend of wind speed becomes small at first, then it is bigger, finally it does not change a lot, the trend of fluctuation changes a lot. The trend of wind speed and fluctuation in cluster 2 stays steady. As to cluster 3, the trend of wind speed changes a lot, it becomes small at first and then becomes bigger and bigger, while the trend of fluctuation stays steady. However, both the trends of wind speed and fluctuation don’t change a lot in cluster 4.
Therefore, the 4 clusters obtained by multiple views are not only different on statistical information, but also different on trends information.

Finally, SVR [20] is used to build models for different clusters. While an unseen sample is assigned to a cluster according to the distances to the center of each cluster, the model on the corresponding cluster is used to predict the value of wind speed. The errors for the one-third test data are listed in TABLE III. We can know that models on cluster 2 and cluster 3 perform well, while models on cluster 1 and 4 perform worse. It also shows that it is necessary to divide the wind speed into different patterns because of the different performance of different clusters.

Table 3. Errors of Wind Speed Prediction in 4 clusters

| NO. | Error | 1      | 2      | 3      | 4      |
|-----|-------|--------|--------|--------|--------|
| MAE |       | 1.7957 | 0.786  | 0.9573 | 1.3511 |
| MSE |       | 5.6466 | 1.265  | 1.7719 | 3.252  |
| Number of training samples | | 3312   | 29562  | 18010  | 4771   |

5. Conclusion

Because of the randomness, fluctuation and uncertainty of wind, great challenges are brought to the wind speed patterns mining under a single angle. In the paper, a method based on multiple views for extracting patterns of wind speed variation is proposed. In the process of clustering, it can get the
latent clusters based both on statistical information and trend information. It can also take each trend as an independent overall feature. The experiment on real wind farm proves that the latent patterns of wind speed variation can be identified effectively based on multiple views. It is found that mean wind speed and fluctuation level are important in the statistical features, and wind speed variation in past states shows different rules in different clusters. Meanwhile, wind speed can be well predicted in the cluster with low wind speed and small fluctuation. What's more, while the wind speed and its fluctuation range are larger, wind speed forecasting model performs worse. It reminds us that different models need to be constructed with different patterns, and range of error in different clusters is different. It can be seen that the multi-view clustering method can better find the wind speed patterns from a data-driven perspective, and help to improve wind speed forecasting accuracy. In future, we will collect more datasets from different wind farms to mining the patterns of wind speed variation, and extract more information of wind speed to carry out the work. In addition, how to effectively build wind speed forecasting model according to different sub-pattern is a valuable work.

6. References
[1] Council, G. W. E. (2013). Global wind statistics 2012. Global Wind Report.
[2] H G. V. Kuik, B. Ummels, and R. Hendriks(2007). Sustainable Energy Technologies. Amsterdam, The Netherlands: Springer.
[3] A Kusiak, A., Zheng, H., & Song, Z. (2009). Wind farm power prediction: a data-mining approach. Wind Energy, 12(3), 275-293.
[4] Saavedra-Moreno, B., Salcedo-Sanz, S., Carro-Calvo, L., Gascón-Moreno, J., Jiménez-Fernandez, S., Prieto, L., 2013. Very fast training neural-computation techniques for real-measure-correlate-predict wind operations in wind farms. J. WindEng. Ind. Aerodyn.116, 49–60.
[5] Mohandes, M. A., Halawani, T. O., Rehman, S., & Hussain, A. A. (2004). Support vector machines for wind speed prediction. Renewable Energy, 29(6), 939-947.
[6] Torres, J. L., Garcia, A., De Blas, M., & De Francisco, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy, 79(1), 65-77.
[7] Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., & Pytharoulis, I., et al. (2008). Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering. Journal of Wind Engineering & Industrial Aerodynamics, 96(12), 2348-2362.
[8] Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion, 21(1), 273-284.
[9] Jiang, X., Jiang, C., Peng, M., Lin, H., & Li, Z. (2010). A short-term combination wind speed forecasting method considering seasonal periodicity and time-continuity. Dianli Xitong Zidonghua(Automation of Electric Power Systems), 34(15), 75-79.
[10] Sanchez, I. (2006). Short-term prediction of wind energy production. International Journal of Forecasting, 22(1), 43-56.
[11] Tascikaraoglu, A., & Uzonoglu, M. (2014). A review of combined approaches for prediction of short-term wind speed and power. Renewable and Sustainable Energy Reviews, 34, 243-254.
[12] Faria, D. L., Castro, R., Philippart, C., & Gusmao, A. (2009, March). Wavelets pre-filtering in wind speed prediction. In Power Engineering, Energy and Electrical Drives, 2009. POWERENG'09. International Conference on (pp. 168-173). IEEE.
[13] Fei, S. W. (2016). A hybrid model of EMD and multiple-kernel RVR algorithm for wind speed prediction. International Journal of Electrical Power & Energy Systems, 78, 910-915.
[14] Ouyang, T., Zha, X., Qin, L., Xiong, Y., & Xia, T. (2016). Wind power prediction method based on regime of switching kernel functions. Journal of Wind Engineering and Industrial Aerodynamics, 153, 26-33.
[15] Wang, C., Lu, Z., & Qiao, Y. (2009, April). Modeling of wind pattern and its application in
wind speed forecasting. In Sustainable Power Generation and Supply, 2009. SUPERGEN’09. International Conference on (pp. 1-6). IEEE.

[16] Kusiak, A., & Li, W. (2010). Short-term prediction of wind power with a clustering approach. Renewable Energy, 35(10), 2362-2369.

[17] Hu, Q., Su, P., Yu, D., & Liu, J. (2014). Pattern-based wind speed prediction based on generalized principal component analysis. IEEE Transactions on Sustainable Energy, 5(3), 866-874.

[18] Lee, H. G., Piao, M., & Shin, Y. H. (2015). Wind Power Pattern Forecasting Based on Projected Clustering and Classification Methods. Etri Journal, 37(2), 283-294.

[19] Wani, M. R., Wani, M. A., & Riyaz, R. (2016, November). Cluster based approach for mining patterns to predict wind speed. In Renewable Energy Research and Applications (ICRERA), 2016 IEEE International Conference on (pp. 1046-1050). IEEE.

[20] Santamaría-Bonfil, G., Reyes-Ballesteros, A., & Gershenson, C. (2016). Wind speed forecasting for wind farms: A method based on support vector regression. Renewable Energy, 85, 790-809.

[21] Shao, W., He, L., Lu, C. T., & Philip, S. Y. (2016, December). Online multi-view clustering with incomplete views. In Big Data (Big Data), 2016 IEEE International Conference on (pp. 1012-1017). IEEE.