Recognition of Wind Speed Patterns Using Multi-Scale Subspace Grids with Decision Trees

M. Arif Wani*, Mehmet Yesilbudak**‡

*Department of Computer and Electrical Engineering and Computer Science, School of Natural Sciences, Mathematics and Engineering, California State University

**Department of Electric and Energy, Vocational College of Haci Bektas Veli, Nevsehir University

awani@csub.edu, myesilbudak@nevsehir.edu.tr

‡Corresponding Author; Department of Electric and Energy, Vocational College of Haci Bektas Veli, Nevsehir University, Nevsehir, Turkey, Tel: +90 384 441 3432, Fax: +90 384 441 3441, myesilbudak@nevsehir.edu.tr

Received: 12.06.2013 Accepted: 15.06.2013

Abstract-The wind speed patterns are essential and indispensable requirement for the efficient utilization of the wind power generated by wind turbines. For this reason, this paper proposes a new approach in order to recognize the wind speed patterns from the multidimensional meteorological data. The meteorological dataset used in this study includes wind direction, air temperature, atmospheric pressure, relative humidity and wind speed parameters. Firstly, the proposed approach eliminated the dimensionality problem of the total dataset by means of obtaining the lower dimensional subspaces with the principal component analysis and the multiple discriminant analysis. Secondly, the proposed approach alleviated the problem of small sample sizes by means of achieving the coarse scales as generic rules at the lower dimensional subspaces. The total dataset includes 3244 observations for each meteorological parameter. In this study, 3100 data points were used for extracting the rules and 144 data points were utilized for testing the extracted rules. As a result, it is mined that the proposed approach leads to reveal the wind speed patterns in a usable and comprehensive manner.

Keywords-Subspace grid-based approach, multi-scale approach, multidimensional meteorological data, wind speed, rule extraction

1. Introduction

Many countries have focused on the research and the installation of renewable energy sources for the purpose of overcoming the challenges caused by the exhaustion of fossil fuels and the increase of climate changes [1]. Renewable energy sources are categorized as hydropower, geothermal, biomass, solar, wind and marine energies [2]. Among them, the wind energy is one of the most promising sources in electricity generation [3]. In wind energy systems, the wind power has an uncontrollable structure due to the non-stationary nature of the wind speed [4]. So, the wind speed parameter plays a critical role in wind power prediction [5]. However, the major problem in wind power prediction is to mine the wind speed patterns from multidimensional data.

Many methods have been developed for the classification of multidimensional data in the literature. A support vector machine model based on recursive feature elimination was utilized for the multi-class categorization of microarray data and the feature selection accuracy was improved in different kernel settings [6]. A data-dependent kernel machine model was proposed for maximizing the separability of the training data and the possible training bias was reduced in the classification of microarray data [7]. A sequential diagonal linear discriminant analysis was implemented for the feature identification in microarray data and the class separation was upgraded by recomputing the model parameters with a robust t-test score [8]. An incremental hybrid approach including principal component analysis and multiple discriminant analysis was presented for the microarray classification and the singular scatter matrix problem caused by small training samples was solved effectively [9]. A subspace grid-based model was designed for the rule extraction in the microarray classification and the extracted rules obtained by the subspace grids with multiple discriminant analysis was more accurate than the ones obtained by the subspace grids with principal component analysis [10]. Similar to these studies, many correlated variables were transformed into a smaller number of uncorrelated variables using the multidimensional wind speed data and the proposed model avoided from the intensive computation time [11]. An evolutionary computing approach was implemented for the dimension reduction and the synoptic pressure patterns were extracted for long-term wind speed estimation [12]. A proper orthogonal decomposition approach was developed for the non-stationary wind speeds and the number of variables was reduced and the complexity of the wind speed data was decreased [13]. Self organizing maps converted the multidimensional feature space to the minimum quantization error and similarity or dissimilarity to the baseline was determined for the wind speed patterns [14]. A multivariate dimension reduction approach was applied to the spatially distributed wind power data and the designed model limited the number of random statistical variables and it prioritized the order of the alternative statistical variables [15].

Unlike the studies in the literature, on the one hand, principal component analysis and multiple discriminant analysis were employed for reducing the multidimensional meteorological data to the lower dimensional subspaces. On the other hand, multi-scale grids were utilized for creating coarse scales at the lower dimensional subspaces. As a result, the dimensionality problem and the issues related to the small sample size have been disposed and many reasonable wind speed patterns were uncovered remarkably in this study. Furthermore, the proposed approach improves the classification accuracy of the multidimensional meteorological data significantly.

2. Meteorological Data Characteristics

A meteorological station placed in Poyracık, Turkey generated the multidimensional meteorological data used in this study by storing wind direction, air temperature, atmospheric pressure, relative humidity and wind speed parameters at 10-min intervals. The units of meteorological parameters were determined as grad (G) for wind direction, Celsius (°C) for air temperature, hectopascal (hPa) for atmospheric pressure, percent (%) for relative humidity and meter per second (m/s) for wind speed. The minimum values of them were observed as 302.991 G, 15.214 °C, 927.137 hPa, 19.746% and 5.022 m/s, respectively. The mean values of them were recorded as 349.694 G, 24.318 °C, 934.444 hPa, 66.647% and 10.234 m/s, respectively. The maximum values of them were measured as 359.984 G, 31.308 °C, 938.906 hPa, 99.987% and 17.501 m/s, respectively. The time-dependent variations belong to the meteorological parameters are illustrated in Fig. 1 and Fig. 2.

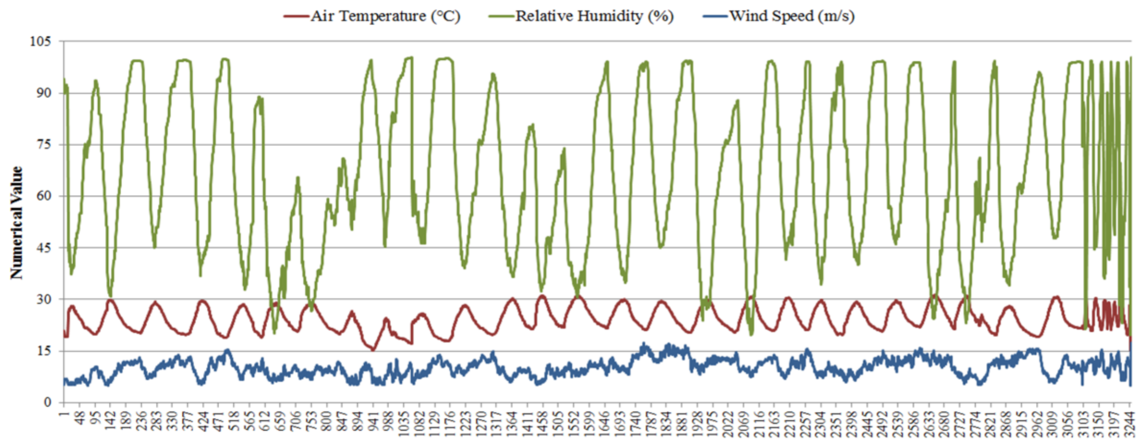

Fig. 1. Air temperature, relative humidity and wind speed values in the total dataset

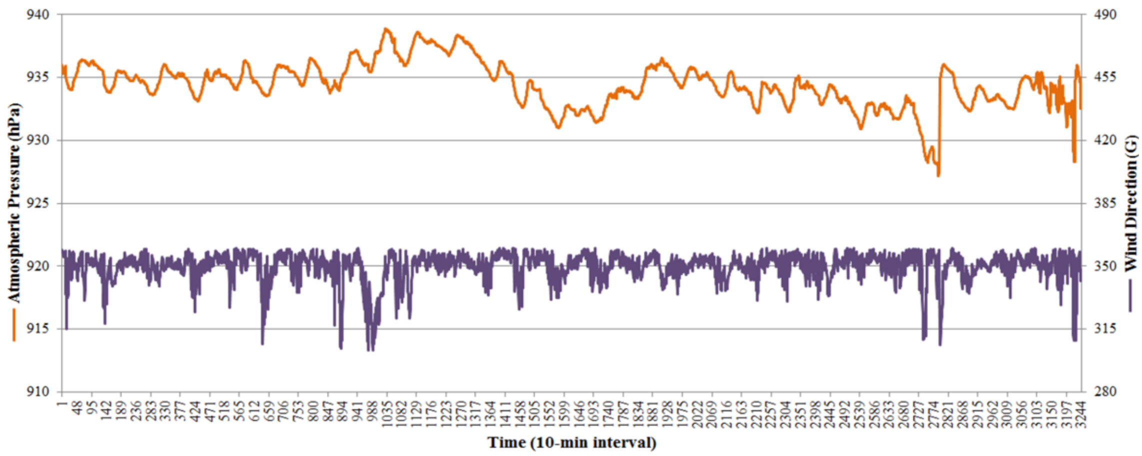

Fig. 2. Atmospheric pressure and wind direction values in the total dataset

3. Multi-Scale Subspace Grid-Based Approach

The dimensionality problem and the issues caused by the small sample size are the main problems encountered in case of recognizing patterns from multidimensional data [16]. Since, the data dimension leads the intensive computations and the small sample size makes the rule extraction difficult [17]. For these reasons, multi-scale subspace grid-based approach is needed for obtaining lower dimensional subspaces and creating coarse and fine scales at lower dimensional subspaces. The features produced by coarse scales represent more generic and stable rules while the features produced by fine scales stand for specific and less consistent rules [18]. A 2-dimensional subspace obtained from the multidimensional data is shown in Figure 3. In this figure, firstly, the related subspace is divided into coarse grids. Secondly, coarse grids containing multi-class data are also divided into fine grids. Thus, the classification accuracy is improved in a good manner.

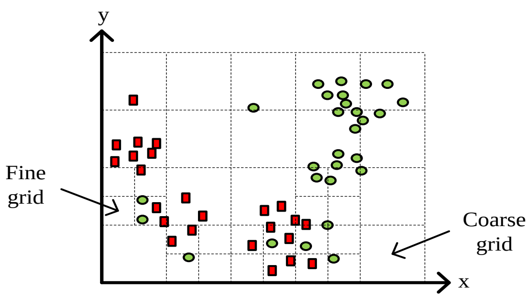

Fig. 3. Conversion multidimensional data to a 2-dimensional space with coarse and fine grids

3.1. Dimensionality Reduction and Grid Construction

The dimensionality problem of multidimensional meteorological data was overcome by means of employing principal component analysis (PCA) and multiple discriminant analysis (MDA) in this study. PCA replaces the original correlated variables with the new uncorrelated variables called principal components by maximizing the variance of original variables [19]. However, MDA generates discriminative components by maximizing the difference between the classes of original variables [20]. The theoretical expressions and the mathematical equations belong to PCA and MDA are described as in [21]. MDA outperforms PCA in case each class in multidimensional data is represented with a single Gaussian distribution and a common covariance matrix. Otherwise, PCA surpasses MDA in case the number of samples for each class in training dataset is limited [22].

Each observation in multidimensional meteorological data was projected along four projection vectors including two projection vectors performed by PCA and two projection vectors comprised of MDA in this study. Three 2-dimensional subspaces were constructed using these four projection vectors. In first one, two vectors from PCA were utilized for spreading the projected data without using class labels. In second one, two vectors from MDA were employed for spreading the projected data using class labels. In both cases, fine grids were constructed. In third one, a vector from PCA and another vector from MDA were used for constructing coarse grids. As a result of these processes, multidimensional meteorological data is converted to three 2-dimensional subspaces containing coarse and fine grids for the purpose of recognizing wind speed patterns in this study.

3.2. Rule Extraction Using Decision Trees

A recursive algorithm based on decision trees was implemented in multi-scale subspace grids for the purpose of extracting rules about the wind speed patterns. Decision trees are utilized for dividing up a large collection of observations into smaller sets of observations by applying a sequence of simple decision rules [23]. In decision trees, each non-leaf node represents a test for an attribute, each branch indicates an outcome for the test and each leaf node denotes a class label [24]. A high-level summary of the recursive-based decision tree approach implemented in this study is expressed below [22].

i. Let S be a set of instances at a node and $S_1, S_2, \ldots, S_n$ be subsets of S,

ii. Let $A_1, A_2, \ldots, A_m$ be a set of m vectors and $C_1, C_2, \ldots, C_p$ be a set of p classes,

iii. Let $Range(A_i)$ be a set of possible values of a vector A_i,

iv. Represent each instance in S with a tuple of $(V_1, V_2, \ldots, V_m, C_k)$ where $V_i \, \varepsilon \, Range(A_i)$ and $C_k \, \varepsilon \, C$,

v. Compute the information entropy of partition process by,

$$\mathrm{I(S_1)} = -\sum \left(\mathrm{P_{S_1,C_k} Log_2 P_{S_1,C_k}}\right)$$

vi. Compute the randomness measure of instance distribution in S_l over the possible classes in C by,

$$\mathrm{E(A_i, S)} = \sum\nolimits_{\mathrm{V_1 \, \varepsilon \, Range(A_i)}} \left(\mathrm{I(S_1)} \frac{\mathrm{n(S_1)}}{\mathrm{n(S)}}\right)$$

vii. Maximize the quantity of $Gain(A_i, S)$ and choose A_i vector as a branch by,

$$\mathrm{Gain(A_i, S) = I(S) - E(A_i, S)}$$

4. Results and Discussion

This research used the dataset where wind speed was characterized by wind direction, air temperature, atmospheric pressure and relative humidity. In consequence of the experimental results, it was observed that only the use of coarse grids is better than using both coarse and fine grids for the reason that fine grids create the specialized rules. For instance, *IF Air Temperature is 19.0 to 20.0 THEN Wind Speed is 5.0* is an extracted rule, but this case is not unique. The same wind speed can be gotten at another air temperature as well. So, the generalization capability of the learning process will be poor in this case. Therefore, the usage of fine grids is not suitable for the application of our interest. Thus, we have only concentrated on the usage of coarse grids. With coarse grids, the training data gave the classification accuracy of 62.0% and the testing data produced the classification accuracy of 75.69%. In case the classification accuracy of the training data was improved to 99.9% by using fine grids, the classification accuracy of the testing data dropped to 32%. This is expected because of the mentioned reasons above. As a result, the average classification accuracy was achieved about 70% in this study.

1: IF Relative Humidity < 46.6060
AND Wind Direction < 340.9863
THEN Wind Speed is (>= 5.0220 & < 9.1817)
2: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction (>= 321.9887 & < 340.9863)
AND Atmospheric Pressure < 931.0600
THEN Wind Speed is (>= 5.0220 & < 9.1817)
3: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction < 321.9887
AND Atmospheric Pressure < 934.9830
THEN Wind Speed is (>= 5.0220 & < 9.1817)
4: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction (>= 321.9887 & < 340.9863)
AND Atmospheric Pressure (< 934.9830 & >= 931.060)
THEN Wind Speed is (>= 9.1817 & < 13.3413)
5: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction < 340.9863
AND Atmospheric Pressure >= 934.9830
THEN Wind Speed is (>= 5.0220 & < 9.1817)
6: IF Relative Humidity < 46.6060
AND Wind Direction >= 340.9863
AND Atmospheric Pressure < 931.0600
THEN Wind Speed is (>= 5.0220 & < 9.1817)
7: IF Relative Humidity < 46.6060
AND Wind Direction >= 340.9863
AND Atmospheric Pressure (>= 931.060 & < 934.983)
AND Air Temperature < 25.9433
THEN Wind Speed is (>= 9.1817 & < 13.3413)
8: IF Relative Humidity < 46.6060
AND Wind Direction >= 340.9863
AND Atmospheric Pressure (>= 931.060 & < 934.983)
AND Air Temperature >= 25.9433
THEN Wind Speed is (>= 5.0220 & < 9.1817)
9: IF Relative Humidity < 46.6060
AND Wind Direction >= 340.9863
AND Atmospheric Pressure >= 934.9830
THEN Wind Speed is (>= 5.0220 & < 9.1817)
10: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction >= 340.9863
AND Atmospheric Pressure < 931.0600
AND Air Temperature < 25.9433
THEN Wind Speed is (>= 5.0220 & < 9.1817)
11: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction >= 340.9863
AND Atmospheric Pressure < 931.0600
AND Air Temperature >= 25.9433
THEN Wind Speed is (>= 9.1817 & < 13.3413)
12: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction >= 340.9863
AND Atmospheric Pressure (>= 931.060 & < 934.983)
THEN Wind Speed is (>= 9.1817 & < 13.3413)
13: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction >= 340.9863
AND Atmospheric Pressure >= 934.9830
AND Air Temperature < 20.5787
THEN Wind Speed is (>= 5.0220 & < 9.1817)
14: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction >= 340.9863
AND Atmospheric Pressure >= 934.9830
AND Air Temperature (>= 20.5787 & < 25.9433)
THEN Wind Speed is (>= 9.1817 & < 13.3413)
15: IF Relative Humidity (>= 46.6060 & < 73.4660)
AND Wind Direction >= 340.9863
AND Atmospheric Pressure >= 934.9830
AND Air Temperature >= 25.9433
THEN Wind Speed is (>= 9.1817 & < 13.3413)
16: IF Relative Humidity >= 73.4660
AND Wind Direction < 340.9863
AND Air Temperature < 20.5787
AND Atmospheric Pressure < 934.9830
THEN Wind Speed is (>= 9.1817 & < 13.3413)
17: IF Relative Humidity >= 73.4660
AND Wind Direction < 340.9863
AND Air Temperature < 20.5787
AND Atmospheric Pressure >= 934.9830
THEN Wind Speed is (>= 5.0220 & < 9.1817)
18: IF Relative Humidity >= 73.4660
AND Wind Direction < 340.9863
AND Air Temperature >= 20.5787
AND Atmospheric Pressure < 934.9830
THEN Wind Speed is (>= 13.3413 & < 17.5010)
19: IF Relative Humidity >= 73.4660
AND Wind Direction < 340.9863
AND Air Temperature >= 20.5787
AND Atmospheric Pressure >= 934.9830
THEN Wind Speed is (>= 9.1817 & < 13.3413)
20: IF Relative Humidity >= 73.4660
AND Wind Direction >= 340.9863
THEN Wind Speed is (>= 9.1817 & < 13.3413)

5. Conclusion

In this study, the wind speed patterns were mined using the multidimensional meteorological data. For this purpose, wind direction, air temperature, atmospheric pressure and relative humidity parameters were employed in multi-scale subspace grids. Firstly, it is uncovered that coarse grids improve the classification accuracy more than fine grids in the recognition of wind speed patterns. Secondly, it is shown that the classification accuracy was obtained as 62.0% for training data and 75.69% for testing data. So, the proposed approach achieved the generalization capability in a good way. In future studies, the multi-scale subspace grid-based approach should also be implemented in the recognition of wind power patterns of wind energy conversion systems.

References

[1] Y.C. Shen, G.T.R. Lin, K.P. Li and B.J.C. Yuan, “An assessment of exploiting renewable energy sources with concerns of policy and technology”, Energy Policy, vol. 38, no. 8, pp. 4604-4616, August 2010.

[2] N.L. Panwar, S.C. Kaushik and S. Kothari, “Role of renewable energy sources in environmental protection: A review”, Renewable and Sustainable Energy Reviews, vol. 15, no. 3, pp. 1513-1524, April 2011.

[3] I. Colak, S. Sagiroglu, M. Demirtas and M. Yesilbudak, “A data mining approach: Analyzing wind speed and insolation period data in Turkey for installations of wind and solar power plants”, Energy Conversion and Management, vol. 65, pp. 185-197, January 2013.

[4] G. Zhang, H.X. Li and M. Gan, "Design a wind speed prediction model using probabilistic fuzzy system", IEEE Transactions on Industrial Informatics, vol. 8, no. 4, pp. 819-827, November 2012.

[5] I. Colak, S. Sagiroglu and M. Yesilbudak, "Data mining and wind power prediction: A literature review", Renewable Energy, vol. 46, pp. 241-247, October 2012.

[6] H. Chai and C. Domeniconi, "An evaluation of gene selection methods for multi-class microarray data classification", 2nd European Workshop on Data Mining and Text Mining in Bioinformatics, pp. 7-14, 24 September 2004, Pisa, Italy.

[7] H. Xiong, Y. Zhang and X.W. Chen, "Data-dependent kernel machines for microarray data classification", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 4, no. 4, pp. 583-595, October-December 2007.

[8] R.P. Regi, A. Ortega and S. Asgharzadeh, "Sequential diagonal linear discriminant analysis (seqdlda) for microarray classification and gene identification", IEEE Computational Systems Bioinformatics Conference, Workshops and Poster Abstracts, pp. 112-113, 8-11 August 2005, Stanford, California.

[9] M.A. Wani, "Incremental hybrid approach for microarray classification", 7th International Conference on Machine Learning and Applications, pp. 514-520, 11-13 December 2008, San Diego, California.

[10] M.A. Wani, "Microarray classification using sub-space grids", 10th International Conference on Machine Learning and Applications, pp. 389-394, 18-21 December 2011, Hawaii, USA.

[11] A. Kusiak and W. Li, "Estimation of wind speed: A data-driven approach", Journal of Wind Engineering and Industrial Aerodynamics, vol. 98, no. 10-11, pp. 559-567, October-November 2010.

[12] L.C. Calvo, S.S. Sanz, N.K. Bossi, A.P. Figueras, L. Prietoc, R.G. Herrera and E.H. Martín, "Extraction of synoptic pressure patterns for long-term wind speed estimation in wind farms using evolutionary computing", Energy, vol. 36, no. 3, pp. 1571-1581, March 2011.

[13] L. Chen and C.W. Letchford, "Proper orthogonal decomposition of two vertical profiles of full-scale nonstationary downburst wind speeds", Journal of Wind Engineering and Industrial Aerodynamics, vol. 93, no. 3, pp. 187-216, March 2005.

[14] E. Lapira, D. Brisset, H.D. Ardakani, D. Siegel and J. Lee, "Wind turbine performance assessment using multi-regime modeling approach", Renewable Energy, vol. 45, pp. 86-95, September 2012.

[15] D.J. Burke and M.J.O. Malley, "A study of principal component analysis applied to spatially distributed wind power", IEEE Transactions on Power Systems, vol. 26, no. 4, pp. 2084-2092, November 2011.

[16] W.H. Yang and D.Q. Dai, "Two-dimensional maximum margin feature extraction for face recognition", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 39, no. 4, pp. 1002-1012, August 2009.

[17] Y. Li, "On incremental and robust subspace learning", Pattern Recognition, vol. 37, no. 7, pp. 1509-1518, July 2004.

[18] G. Sundaramoorthi, A. Yezzi and A.C. Mennucci, "Coarse-to-fine segmentation and tracking using sobolev active contours", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 5, pp. 851-864, May 2008.

[19] A. Pascual, M.L. Martín, F. Valero, M.Y. Luna and A. Morata, "Wintertime connections between extreme wind patterns in Spain and large-scale geopotential height field", Atmospheric Research, vol. 122, pp. 213-228, March 2013.

[20] S.H. Fang and T.N. Lin, "Projection-based location system via multiple discriminant analysis in wireless local area networks", IEEE Transactions on Vehicular Technology, vol. 58, no. 9, November 2009.

[21] J. McBain and M. Timusk, "Feature extraction for novelty detection as applied to fault detection in machinery", Pattern Recognition Letters, vol. 32, no. 7, pp. 1054-1061, May 2011.

[22] M.A. Wani, "Introducing subspace grids to recognise patterns in multidimensional data", 11th International Conference on Machine Learning and Applications, pp. 33-39, 12-15 December 2012, Florida, USA.

[23] M.J.A. Berry and G. Linoff, "Decision trees", Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Indianapolis: Wiley Publishing, 2004, pp. 165-166.

[24] J. Han and M. Kamber, "Classification by decision tree induction", Data Mining: Concepts and Techniques, San Francisco: Morgan Kaufmann Publishers, 2006, pp. 291-292.