Short Term Estimation and Analysis of Wind Speed using Data Mining Techniques

P. Sardar Maran
Research Scholar
Centre for Earth & Atmospheric Sciences
Sathyabama University, Chennai

R.Ponnusamy, Ph.D
Principal
Madha Engineering College Kundrathur
Chennai

ABSTRACT

Weather Data Mining is a form of Data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. In this paper we used meteorological data mining to analyze wind speed behavior. The two years data was recorded from 2010 - 2011 daily historical data by 50m instrumented meteorological station at Sathyabama University. After preprocessing the data, we applied data mining techniques: Association rules, Classification, and Cluster analysis. From these tasks, we found the most appropriate of these techniques to be applied on weather data is classification task, especially the neural networks method because the nature of the data is time series.

Keywords

Data mining, Wind Speed, Association rules, Classification, Cluster

1. INTRODUCTION

Weather prediction has been one of the most interesting and fascinating domain. It plays an important role in natural and human life.(1) Wind energy is one of the most promising alternative energy technologies of the future. Throughout recent years, the amount of energy produced by wind-driven turbines has increased exponentially due to significant breakthroughs in turbine technologies, making wind power economically compatible with conventional sources of energy [2]. Due to increasing demand for renewable energy resources, wind energy and its associated issues have received more attention recently.[3,4,5] The power-generating efficiency of a wind turbine can be significantly increased if the turbine's operation is controlled based on the information of wind and wind changes at the turbine location.[2] There are various methods in soft computing techniques are available for the predication of wind speed namely Artificial Neural Network, Numerical Weather Prediction and Fuzzy Logic.[6,7,8]

India has the fifth largest installed wind power capacity in the world. Wind power accounts for 6% of India's total installed power capacity, and it generates 1.6% of the country's power. It is estimated that 6GW of additional wind power capacity will be installed in India by 2012, taking the total installed capacity beyond 15GW. (Indian Wind Energy Summit 2012).

This paper investigates the wind speed prediction domain of data mining using a case study. It showed what kind of data could be collected, how could we preprocess the data, how to apply data mining methods on the data, and finally how can we benefited from the discovered knowledge. There are many kinds of knowledge can be discovered from data. In this work we investigated the most common ones which are association, classification, clustering and outlier detection.

With increasing agricultural and industrial activities in the country, the demand for energy is also increasing, K. Sreelakshmi and P. Ramakanthkumar [2] introduced a feedforward neural network model for short term wind speed prediction, which uses back propagation algorithm. They utilize six different parameters values (Mean temperature, Humidity, Wind gust, Wind direction, Barometric pressure and Wind speed) as the input to model. Finally, they reach on model that predicted wind speed differs from the actual value by max 5%.

The paper is structured as follows: in Section 2 summaries related works in weather data mining. Section 3 gives a general description of the data we used in our case study. Section 4 describes the preprocess stage of the used data. Section 5 illustrates our experiments about applying data mining methods on the wind speed data. Finally we conclude this paper with a summary.

2. AREA OF STUDY

In this study the date we have collected from 50metre instrumented Meteorological Tower located at Sathyabama University, near the east coast of Chennai, Latitude 12.9 N, and Longitude 80.0 E. The 50m meteorological tower is having four levels of Temperature and Humidity sensors at the heights of 2m, 8m, 16m and 50m.The wind speed and wind direction sensors at five levels at the heights of 2m, 8m,16m,32m and 50m. The Atmospheric Pressure is at the ground level. The observed data contains 3 years and 6 months cover the years from January 2010 to July 2013 daily ten minutes data. Total records of 150957 rows selected for this study. Five different weather parameters data such as average temperature, pressure, humidity, wind direction and wind speed involves in this study as shown in the table 1.

Table 1. Weather Parameters

PARAMETERS	SYMBOL
Average Temperature	T-Avg
Pressure	Pres
Relative Humidity	Rh-Avg
Wind Direction	WD
Wind Speed	WS

3. DATA PREPARATION

The collected data was preprocessed for better input for the data mining technique. Then the data is integrated and rearranged in the log format. The lag values of wind speed for today, yesterday and day before yesterday are calculated. The wind speed attributes are classified in to five group's namely very high, high, middle, low and very low. The subsets of attributes are average temperature, pressure, humidity and wind direction. The normalized values of wind speed are converted into zero and one. The following Table 2 shows the as the normalized values.

Table 2. Normalized Weather Parameters

Windspeed	Temp Avg	Hum Avg	Pres	WD	WSlog1	WSlog2	WSlog3	Day	Month
very low	0.741	0.77	0.957	0.287	0.651	0.685	0.642	25	1
very low	0.25	0.79	0.859	0.295	0.685	0.642	0.685	26	1
very low	0.453	0.82	0.835	0.982	0.642	0.685	0.651	27	1
high	0.515	0.811	0.758	0.952	0.885	0.851	0.881	28	1
high	0.857	0.781	0.658	0.985	0.851	0.881	0.811	29	1
high	0.83	0.711	0.698	0.955	0.881	0.811	0.685	30	1
low	0.422	0.681	0.655	0.9	0.611	0.685	0.662	31	1
low	0.539	0.81	0.698	0.9	0.685	0.662	0.6	1	1
low	0.436	0.812	0.708	0.972	0.662	0.6	0.685	2	1
low	0.471	0.81	0.7	0.942	0.6	0.685	0.811	3	1
middle	0.091	0.861	0.798	0.254	0.685	0.811	0.685	4	1
middle	0.12	0.72	0.798	0.985	0.811	0.685	0.625	5	1
middle	0.07	0.71	0.835	0.265	0.785	0.725	0.765	6	1
middle	0.08	0.71	0.835	0.222	0.725	0.765	0.751	7	1
middle	0.52	0.81	0.8	0.212	0.765	0.751	0.785	8	1
high	0.371	0.721	0.753	0.3	0.851	0.985	0.851	9	2
high	0.921	0.711	0.605	0.298	0.985	0.851	0.732	10	2
middle	0.72	0.651	0.698	0.201	0.851	0.732	0.685	11	2
low	0.628	0.745	0.698	0.2	0.632	0.685	0.671	12	2
low	0.464	0.784	0.6	0.972	0.685	0.671	0.621	13	2
low	0.467	0.861	0.698	0.962	0.671	0.621	0.662	14	2
low	0.38	0.72	0.658	0.952	0.621	0.662	0.685	15	2
low	0.131	0.715	0.753	0.985	0.662	0.685	0.513	16	2
very low	0.836	0.718	0.75	0.955	0.585	0.513	0.511	17	3
very low	0.827	0.701	0.8	0.9	0.513	0.511	0.671	18	3
very low	0.222	0.705	0.859	0.9	0.511	0.671	0.822	19	3
high	0.423	0.825	0.835	0.972	0.671	0.822	0.851	20	3
high	0.486	0.81	0.75	0.942	0.822	0.851	0.693	21	3
high	1.321	0.713	0.698	0.952	0.851	0.693	0.832	22	3
high	1.25	0.71	0.698	0.952	0.693	0.832	0.821	23	4
low	0.682	0.671	0.698	0.952	0.832	0.821	0.685	24	4
low	0.58	0.811	0.6	0.952	0.621	0.685	0.685	25	4
low	0.537	0.812	0.698	0.887	0.685	0.685	0.671	26	4
very low	0.523	0.771	0.7	0.885	0.685	0.571	0.552	27	4
low	0.122	0.841	0.753	0.922	0.671	0.652	0.671	28	4
low	0.851	0.721	0.753	0.989	0.652	0.685	0.822	29	4

After preprocessing the data is visualized using the Weka software as shown in Figure.1

4. DATA MINING

Data mining as an interdisciplinary field draws from statistical analysis, database systems, machine learning, pattern recognition, neural networks, fuzzy systems and other 'soft computing' techniques. Although data mining is young interdisciplinary filed, its methods are quite developed and many of them are practically applicable. Data mining can be defined as a process of discovering new, interesting knowledge, such as patterns, associations, rules, changes, anomalies and significant structures from large amounts of data stored in data banks and other information repositories. It is currently regarded as the key element of a much more elaborate process called Knowledge Discovery in Databases

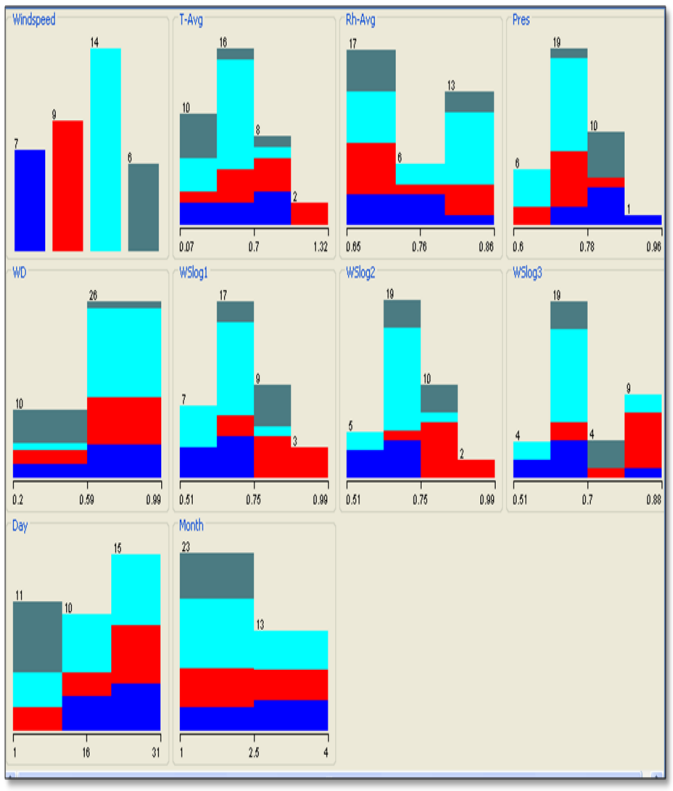

Figure 1. Visualization of Weather Parameters

(KDD). In general, a knowledge discovery process consists of an iterative sequence of the following steps

1.data selection, where data relevant to analysis task are retrieved from database;

2. *data cleaning*, which handles noisy, erogenous, missing or irrelevant data;

3. *data integration (enrichment)*, where multiple heterogeneous data may be integrated into one;

4. *data transformation (coding)*, where data are transformed or consolidated into forms appropriate for different mining algorithms;

5. *data mining*, which is an essential process where intelligent methods are applied in order to extract hidden and valuable knowledge from data;

6. *knowledge representation*, where visualisation and knowledge representation techniques are used to present the mined knowledge to the user.

5. DATA MINING IN WIND SPEED ANALYSIS

In the next sections, we describe the results of applying data mining tasks on our data for each of the four tasks.

6. ASSOCIATION RULES

Association rule mining, playing a critical role in the field of data mining, searches for interesting relationships among items in a given data set. It studies the frequency of items occurring together in transactional databases, and based on a threshold called *support*, identifies the frequent item sets. Another threshold, *confidence* which is conditional probability than an item appears in a transaction when another item appears, is used to pinpoint association rules. For association rules mining, we used FP Growth mining algorithm with min support=0.95 and min confidence= 0.8

It allows finding rules of the form If [wind_sp1 = low, RH_avg = high and T_avg = high] then [wind_sp0 = low where wind_sp1, RH_avg and T_avg] as first rule. However, the previous rule says: when the WSlog2 of wind speed is

low, the average of humidity is high and the average of temperature is high then we found WSlag1 of wind speed is low. The fourth rule is [wind_sp2 = low,wind_sp1 = low, output (windspeed) = low, RH_avg = high, T_avg = high] then [wind_sp0 = low]. That means when the WSlog2 of wind speed is low, WSlog1 of wind speed is low, output of wind speed is low, the average of humidity is high and the average of temperature is high then we found WSlog1 of wind speed is low. The strength of some association rules can be obtained with change min confidence= 0.987

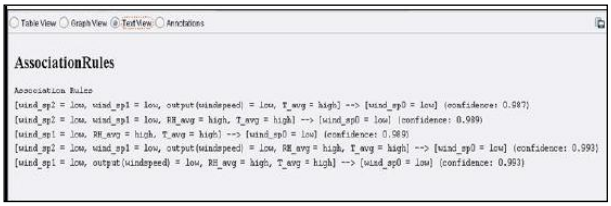

Figure 2. Association Rules

7. CLASSIFICATION

Classification analysis is a data mining task that organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. In wind speed data mining, we classify the degree of wind speed for each day in to one of our classes (very high, high, middle, low and very low), also we spilt data to 70% for training and 30% for testing. In our case study we used three classification methods:

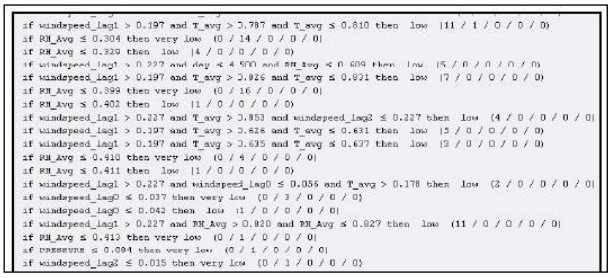

Figure 3. Rule model for wind speed

8. CLUSTERING

Clustering is the organization of data in classes. However, unlike classification, in clustering, class labels are unknown and it is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. In wind speed data mining, clustering has been used to group wind speed state for each day according to changeability of weather for each day. In our case we used KMeans Algorithm to cluster the given data into five groups (k=5) and guide them based on their behavior.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
Day	15.722	16.475	15.782	15.271	21
Month	6.337	7.040	5.868	6.617	2
T_Avg	17.189	20.200	21.809	23.556	16.230
RH_Avg	61.352	63.426	69.277	69.574	82.700
PRESSURE	13.179	12.911	9.589	9.917	15.600
wind_SP_la	158.192	44.716	249.218	322.623	1803
wind_SP_la	10.424	10.613	9.909	9.651	10
wind_SP_la	10.431	10.269	9.907	9.918	2
wind_speed	9.647	9.448	10.587	10.220	7
Wind_Dir	8.584	8.766	11.453	10.387	6.555

Figure 4. Clustering of Wind Speed

9. CONCLUSION

In this paper, we applied data mining tasks (association rules, classification, cluster and outlier analysis) on wind speed data set. We applied association rules by used FPGrowth mining algorithm with min support=0.95 and min confidence= 0.8 and tried to obtain five strength rules with change min confidence to 0.987.

10. REFERENCES

[1]Marwa F. AlRoby, Alaa M. ElHalees, Data Mining Techniques for Wind Speed Analysis,Journal of Computer Engineering Vol 2, No 1 (2011)

[2]Sreelakshmi K, Ramkanthakumar, P., 2008. "Neural Networks for shortterm wind speed prediction". World Academy of Science, Engineering and Technology 42, 721725.

[3]Mrs. C. Beulah Christalin Latha, Dr. (Mrs.) Sujni Paul, Dr.E.Kirubakaran, Mr. Sathianarayanan, Int. J. of Advanced Networking and Applications 608 Volume: 02, Issue:02, Pages:608-613 (2010) A Service Oriented Architecture for Weather Forecasting Using Data Mining

[4]Folorunsho Olaiya, Adesesan Barnabas Adeyemo, Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, I.J. Information Engineering and Electronic Business, 2012, 1, 51-59 Published Online February 2012 in MECS

[5]Muhammad Shaheen, Muhammad Shahbaz, Khalid Afsar Khan Jadoon, Data Mining For Wind Energy Site Selection Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I WCECS 2012, October 24-26, 2012, San Francisco, USA

[6]Lionel Fugon , J_er_emie Juban and George Kariniotakis,European Wind Energy Conference - Brussels, Belgium, April 2008, Data mining for wind power forecasting

[7]Andrew Kusiak*, Haiyang Zheng and Zhe Song, Wind Farm Power Prediction: A Data-Mining Approach, WIND ENERGY Wind Energ. 2009; 12:275–293 Published online 24 September 2008 in Wiley Interscience (www.interscience.wiley.com) DOI: 10.1002/we.295

[8]Sarah N. Kohail, Alaa M. El-Halees Implementation of Data Mining Techniques for Meteorological Data Analysis Volume 1 No. 3, July 2011 ISSN-2223-4985 International Journal of Information and Communication Technology Research

[9]Godfrey C. Onwubolu1, Petr Buryan2, Sitaram Garimella3, Visagaperuman Ramachandran4, Viti Buadromo5and Ajith Abraham6 IADIS European Conference Data Ming 2007, SELF-ORGANIZING DATA MINING FOR WEATHER FORECASTING,ISBN: 978-972-8924-40-9 © 2007 IADIS