EasyChair Preprint

№ 2732

# Wind Power Forecasting Using K-Means Clustering and Convolutional Neural Network

M. Bhavani, S. Mohith Vasan, S. Muthu Kumar and
N. Padma Gokul



February 21, 2020

# WIND POWER FORECASTING USING K-MEANS CLUSTERING AND CONVOLUTIONAL NEURAL NETWORK


Ms.M.Bhavani,

Assistant Professor, Department of computer Science, Rajalakshmi Engineering College.

S.Mohith Vasan,

Department of Computer Science, Rajalakshmi Engineering College

S.Muthu Kumar,

Department of Computer Science, Rajalakshmi Engineering College.

N.Padma Gokul,

Department of Computer Science, Rajalakshmi Engineering College.



## ABSTRACT

Wind energy is a renewable energy that is free, clean and readily available. They also cause less impact to the environment. However, the random nature of the wind and its associated uncertainty create challenges in dealing with the generation of wind power effectively, which can result in unnecessary restrictions to the potential of a wind farm. Wind power forecasting is essential for energy trading and can also be used to decide on economic dispatch based on the forecast. This system makes use of data mining techniques and deep learning to predict the wind power by the combined approach of K-Means Clustering and Convolutional Neural Network. At first, the dataset is preprocessed in MATLAB using Self Organizing Map for dimensionality reduction. Then K-Means is used for clustering the datasets based on meteorological conditions, historical power data and wind turbine parameters. Finally, the cluster is used for training and testing the Convolutional Neural Network to generate the result.



Keywords—Data Mining, K-Means Clustering, Convolutional Neural networks, Self Organizing Map, Deep learning, MATLAB.


## I. INTRODUCTION

Data mining is the process of exploring patterns that are usually large in size are done by methods combining the concept of machine learning, statistics, database systems. There are various techniques that make data mining possible. Self-organizing map is also called as Self-organizing feature map which is used to produce a map that is trained using unsupervised learning. For preprocessing SOM is used to test and train the model also the system makes use of the clustering technique. Clustering involves creating a group of similar data. This is useful in finding the similarities and differences between the data. K-Means is one of the clustering algorithms that work based on partition clustering. It works in an iterative way by partitioning similar data by assigning the data point to the nearest cluster by calculating the distance

between the data point and centroid of the cluster. This will be useful to cluster similar days and to divide this cluster for training and testing purposes.

Deep learning is the subset of machine learning that uses Artificial Neural Networks and different algorithm to learn from input data through experience. It works by performing a task repeatedly each time adjusting it slightly to improve the result. The Neural Networks that we use in deep learning is inspired by the human brain's neural structure. Since the amount of data increasing day by day, deep learning capabilities also keeps growing due to the fact that deep learning algorithms require a large amount of data to learn from. Deep learning allows machine to solve complex problems even when the data is diverse and unstructured. The more they train, better they perform.

The deep learning algorithm that is going to be used here is Convolutional Neural Network or CNN for short. CNN is widely used in image processing, object recognition and video processing. However, research has shown that using CNN for time series forecasting has various advantages over other Neural Networks. Convolutional Neural Network consists of input and output layer with multiple hidden layers, these hidden layers convolve with a dot product. It uses Relu function as an activation function and also includes different layers added to convolution (pooling layers, fully connected layers and normalization layers). Firstly, in traditional Neural Networks, feature extraction and selection need to be built manually. This process is usually time consuming and difficult. In CNN feature extraction is done automatically. Secondly, CNN is more efficient in terms of memory due to its weight sharing feature. Therefore we combine Convolutional Neural Network with K-Means clustering to achieve better accuracy and scalability.

## II. REVIEW OF LITERATURE

Wenbin Wu and Mugen Peng [1] proposed a data mining approach for wind power forecasting has been proposed, which uses of the K-means clustering method and bagging Neural network. The historical data are clustered according to the meteorological conditions and historical power. Pearson correlation coefficient is used to calculate the distance between the forecasting day and the clusters. The input variables of the Neural Network are selected by Relief algorithm to reduce the complexity and the Bagging algorithm is applied to optimize the stability and accuracy of the BPNN. The RMSE and MAE result shows that the proposals have significant efficient.

Sana Mujeeb and Turki Ali Alghamdi [2] proposed an efficient deep learning based prediction model is proposed for wind power forecasting. In this proposed model, there are two phases. In the first phase, WPT is used to decompose the past wind power signals. Other than decomposition the signals and lagged wind power, multiple exogenous inputs such as, calendar variable and Numerical Weather Prediction (NWP) are also given as input to forecast wind power. In the second phase, a new prediction model, Efficient Deep Convolution Neural Network (EDCNN), is used to forecast wind power.

Jinli Dou a and Chun Liu b [3] proposes a short-term wind power forecasting method based on CNN and NWP data in a variety of weather elements like time-dependent, you can easily increase the efficiency of the model using data mining. The CNN can effectively extract the correlation between the weather elements like Temperature, wind speed, humidity over a period of time, so that the wind power prediction can be carried out more accurately. It also has important significance for the future deep learning algorithms.

Jorge Enrique Zafra and Robinson Jimenez Moreno [4] presented the development of two algorithms of Neural networks, which allow the recognition of traffic signs from the realization of a training of Neural networks through Back-Propagation Neural network and another training using CNN, in order to compare them and to know which of the two methods is more efficient in terms of execution time and precision, with respect to the same amount of input images are given as input. Once the networks were trained, random-image tests were performed achieving 98.33% accuracy for the Back-Propagation algorithm and 94.44% accuracy with CNN.

Dr.G.Anandharaj and S.Meena [5] Discuss Medium term wind power prediction based wind speed, wind direction, air temperature and power utilizing wind turbine information collected at 72‟s intervals. A time-series model display way to deal wind behaviour is studied. Both exponential smoothing and data-driven models are created for wind prediction. An overview of the ongoing improvements in machine learning techniques focused on prediction using Ensemble Learning. In this paper, to deal with the training samples dynamics and improve the forecasting accuracy, K-means clustering is utilized to classify the samples into several categories, which contain the data of meteorological conditions and historical power data.

Jinqiang Liu and Xiaoru Wang [6] proposes a novel hybrid methodology for short-term wind power forecasting, successfully combining three individual forecasting models using the Adaptive Neuro Fuzzy Inference System (ANFIS). The Back-Propagation Neural network (BPNN), Radial Basis Function Neural Network (RBFNN), and Least Squares Support Vector Machine (LSSVM) are selected as the individual forecasting models. Results obtained show the advancement of the PCC based data pre-processing method. This paper shows a significant improvement in accuracy with respect to three individual models.

K.P. Moustris and D. Zafirakis [7] proposed forecasting methodology shows that is able to give sufficient and adequate prognosis of wind prediction by a wind turbine in a specific location 8 hours in advance. This will be a useful tool for the operator of a RES system in order to achieve a better monitoring and a better management of the whole system. Artificial Neural Networks (ANN) modelling is applied and appropriate training of the developed ANN models hourly meteorological data is used.

Qianyao Xu and Dawei He [8] proposed a short-term wind power forecasting approach by mining the bad data of Numerical Weather Prediction (NWP). Today’s short-term Wind Power Forecast (WPF) highly depends on the NWP, which contributes the most in the WPF error. This paper first introduces a bad data analyzer to fully study the relationship between the

WPF error with several new extracted features from the raw NWP. Second, a hierarchical structure is proposed, which is composed of a K-means clustering-based bad data detection module and a Neural Network (NN)-based forecasting module.

## III. EXISTING SYSTEM

Wind power forecasting is usually done using 4 approaches. The popular approach is by using Numerical Weather Prediction (NWP) since it is directly linked to weather conditions. It uses mathematical models of oceans and atmosphere to predict the weather conditions. The second method is physical approach where theoretical power curve is used. The third method is statistical approach, which is becoming popular these days due to the increasing popularity of Artificial Intelligence (AI). There are also several time series models to forecast the wind energy. Much research work has shown that the time series models can obtain good forecasting accuracy only for very short-term scenarios. The above literature survey shows wind power forecasting done with various statistical methods. They were surveyed for several factors such as performance efficiency and time efficiency. The proposed system has been designed based on the survey taken.

## IV. PROPOSED SYSTEM

Some of the deep learning algorithms such as RNN shows high performance in wind power forecasting. However, a combined method of CNN and K-Means for wind power forecasting is proposed. Firstly, the dataset is preprocessed by data normalization which is used for structuring a relational database to reduce data redundancy and improve data integrity. The data in the dataset will not be evenly distributed. To construct a model on this dataset without clustering, it could lead to overfitting problems. In order to avoid overfitting, make use of clustering technique to cluster similar days together. Then these clusters are divided into training and testing data. With the training data, construct a CNN model and train it to form the forecasting method. Then the model is tested with testing data and the result is generated. The generated result is evaluated and the performance is calculated.

A sample of data is taken from a large dataset, divided it for training data and testing data, this process is done with help of Self Organizing Map(SOM) using matlab.

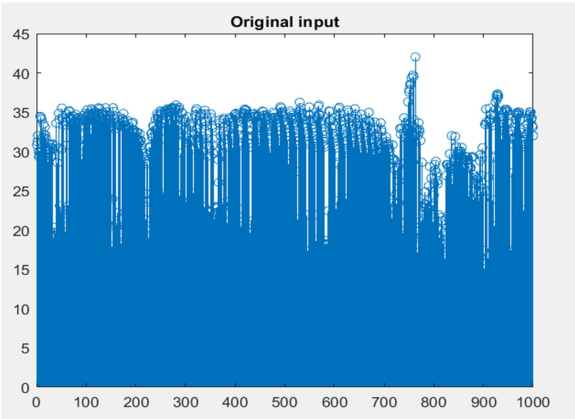

Figure 1: Raw input

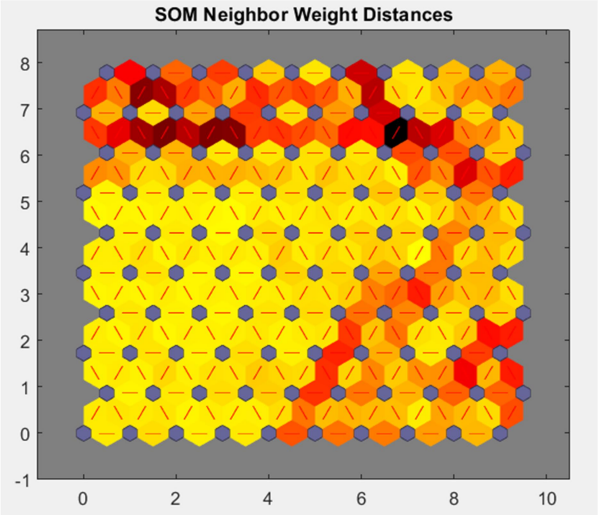

Figure 2: SOM Weight

## IV. CONCLUSION AND FUTURE WORKS

In this paper, a method to forecast electricity loads using joint CNN and K-Means algorithm is proposed. K-Means algorithm is used to large data sets into subsets for constructing models faster and more efficiently. The model is evaluated on a large wind farm power production data set. The dataset is preprocessed in MATLAB using Self Organizing Map for dimensionality reduction.Then CNN is used for classification process. A CNN model is trained and tested, final predicted data is analysed.

For the future work, CNN forecasting abilities by introducing more related factors to training models. In addition, more accuracy can be brought using other clustering algorithms like GMM.

## REFERENCES

[1] Wenbin Wu, Mugen Peng, "A Data Mining Approach Combining K-Means Clustering with Bagging Neural Network for Short-term Wind Power Forecasting" ,arXiv:1709.05261v1 [cs.CE] ,14 Sep 2017.

[2] Sana Mujeeb, Turki Ali Alghamdi, Sameeh Ullah, "Exploiting Deep Learning for Wind Power Forecasting Based on Big Data AnalyticsSmart", 18 October 2019.

[3] Jinli Dou a, Chun Liu b, Bo Wang c, "Short-term Wind Power Forecasting Based on Convolutional Neural Networks",*IOP* Conf. Ser.: Earth Environ. Sci. 170 042023, Jinli Dou *et al* 2018.

[4] Jorge Enrique Zafra, Robinson Jimenez Moreno, Ruben Dario Hernández, "Comparison between Backpropagation and CNN for the Recognition of Traffic Signs", ISSN 0973-4562 Volume 12, Number 17 2017.

[5] Dr.G.Anandharaj, S.Meena, "Improving Ensemble Learning in Ultra Medium-Term Wind Power Prediction Based On Machine Learning Techniques" ,Vol. 08, Issue 8, ||V (IV) || 32-39, August 2018.

[6] Jinqiang Liu, Xiaoru Wang, Yun Lu A, "Novel Hybrid Methodology for Short-term Wind Power Forecasting Based on Adaptive Neuro-Fuzzy Inference System", c 10.1016/j.renene.2016.10.074 ,30 October 2016.

[7] K.P. Moustris, D. Zafirakis, K.A. Kavvadias, J.K. Kaldellis, "Wind Power Forecasting using Historical data and Artificial Neural Networks Modeling" ,2017

[8] Qianyao Xu, Dawei He, Ning Zhang, "A Short-Term Wind Power Forecasting Approach With Adjustment of Numerical Weather Prediction Input by Data Mining", April 23, 2015. TSTE-00676-2014.

[9] Q. Xu, N. Zhang, J. Huang, "A shortterm wind power forecasting approach with adjustment of numberrical weather prediction input by data mining", IEEE Trans. Sustain. Energy, vol. 6, no. 4, pp. 1283-1291, Oct. 2015.

[10] P. Kunzemann, G. Jacobs, and R. Schelenz, "Application of CPS Within Wind Energy-Current Implementation and Future Potential in Industrial Internet of Things", Switzerland: Springer, pp.647-670, 2017.

[11] Zhao Y.N, Ye L, Li Z,"A novel bidirectional mechanism based on time series model for wind power forecasting". 177, 793–803, Appl. Energy 2016.
[12] Tomporowski A, Flizikowski J, Kasner , "Environmental control of wind power technology", Rocznik Ochrona Srodowiska,
[13] , 19, 694–714, 2017.
[14] Athari, M.H.Wang Z., "Impacts of Wind Power Uncertainty on Grid Vulnerability to Cascading Overload Failures". , 9, 128–137 IEEE Trans. Sustain. Energy 2018.
[15] Mujeeb S, Javaid N, Gul H, "Wind Power Forecasting Based on Efficient Deep Convolution Neural Networks. In Proceedings of the 3PGCIC Conference", Antwerp, Belgium, 7–9 November 2019.
[16] Wang J, Niu T, Lu H "A Novel Framework of Reservoir Computing for Deterministic and Probabilistic Wind Power Forecasting". IEEE Trans. Sustain. Energy 2019.
[17] Lin Y, Yang, Wan C, "A Multi-model combination approach for probabilistic wind power forecasting", 10, 226–237, IEEE Trans. Sustain. Energy 2019.