

## Feature Selection used for Wind Speed Forecasting with Data Driven Approaches

Senthil Kumar P\* and Daphne Lopez

School of Information Technology & Engineering, VIT University, Vellore, India.

Received 18 November 2014; Accepted 11 December 2015

### Abstract

Wind speed forecasting is important for wind power generation and integration. In this paper, Nonlinear Autoregressive model process with exogenous input (NARX) is proposed for wind speed forecast. The main aim of this experiment is to forecast wind speed with meteorological time series data as input variable using NARX model. Prior to forecasting, ReliefF feature selection is used for identification of important features for wind speed forecast and reduces the complexity of the model. Performance is evaluated in terms of mean square error when using the feature selection method with the NARX model.

*Keywords:* Wind speed, Feature selection, NARX, ReliefF

### 1. Introduction

Wind energy is one of the important renewable energy sources, and it is an alternative to conventional power generated from fossil fuel. Wind power generation is rapidly increasing and it is a significant contributor to the electricity grid. In recent years, with the large scale expansion of wind farm, more energy is obtained from wind sources. Wind power satisfies the electrical demand and maximizes both economic and environmental benefits. Wind speed forecasting is an important criterion to address the variability and uncertainty of wind speed. The wind speed forecast for the wind energy sector is important due to following reasons like wind power unit commitment decision, load increment and decrement decision, maintenance scheduling and energy storage optimization.

Wind speed is a time series data measured at different intervals of time. A system characterized by time horizon can be used to predict the wind speed. In general wind power sector, forecasting can be segregated into three time horizons as very short term, short term and medium term or long term. A very short term time horizon is for very few hours only. These forecasts are useful for unit commitment and trading in intraday markets. Short term time horizon could be for 12 hours to 48 or 72 hours. These forecasts are useful for trading in the day-ahead market and maintenance scheduling. A medium term time horizon is for a few days (3 to 7 days). These forecasts are useful for maintenance planning of conventional power generation plants [1], [2].

The paper is organized as follows. Section 2 presents review on wind speed forecasting approaches. In section 3 significance of feature selection is discussed and ReliefF feature selection is addressed to select important features for

wind speed forecasting. In section 4 introduction about Artificial Neural Network (ANN) and NARX neural networks model for wind speed forecasting is tested. Finally, section 5 shows result analysis and performance evaluation of wind speed. The framework of the proposed wind speed architecture is shown in Fig.1.

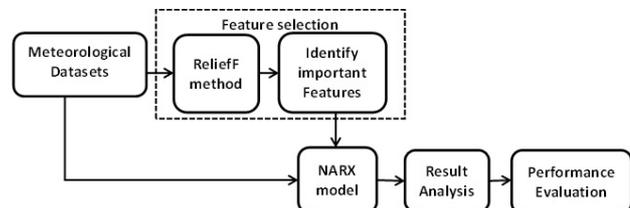


Fig. 1. Framework of wind speed forecasting

### 2. Review on wind speed forecasting approaches

Hui Liu et al. [3] propose a statistical method to predict wind speed and wind power. Forecasting method based on the wavelet and the Improved Time Series a Method (ITSM). Decomposition of the original time series into a number of different sub-series performed by wavelet. ITSM is applied to build the prediction models for every sub-series, and make multi-step forecasting in each sub-series. Calculation of aggregates of sub-series is done to attain the final forecasting for original time series. Forecasting results of the proposed method are compared using classical time series method and BP network methods. The proposed method improves the accuracy of forecasting over the other methods. K.GnanaSheela et al. [4] propose a neural Network based hybrid computing model for wind speed prediction. The proposed hybrid model is combination of Self Organizing feature Maps (SOM) and Multilayer Perceptron (MLP) network that ensures a more reliable wind speed prediction. SOM is used for clustering of input data for determining the relationship between input vectors and MLP for prediction of data. The hybrid model is the improved accuracy with

\* E-mail address: senbe@rediffmail.com

minimal error. S. Salcedo-Sanz et al. [5] proposes short-term wind speed prediction based on a Coral Reefs Optimization algorithm (CRO) and an Extreme Learning Machine (ELM). The approach is based on a feature selection with the CRO, and reduced number of predictive variables out of the total available from the Weather Research and Forecast model (WRF). This set of features used as input of an ELM and provides the wind speed prediction. This paper focus that feature selection process performed with the CRO is robust, and the features selected are very important in the wind speed prediction. Hui Liu et al. [6] establish that hybrid model ARIMA-Artificial Neural Networks (ANN) and ARIMA-Kalman Filter (KF) used for wind speed prediction. Build an ANN forecasting model of a section of wind speed sample. First, an ANN modelling process, a time series ARIMA is used to decide its best structure. Second, Establish a Kalman forecasting model for the same wind speed data. To get the best initial parameters of the Kalman model, a time series ARIMA is employed. Two hybrid models including the ARIMA-ANN, ARIMA Kalman filter and Pure ARIMA are used for multi-step ahead prediction of wind speed data. The performance of the two hybrid models is better than that of the pure ARIMA model, and the performance of the ARIMA-Kalman model is better than that of the ARIMA-ANN model. Finally the two hybrid models have good forecasting accuracy.

### 3. Significance of Feature Selection

A fundamental problem of machine learning is to approximate the functional relationship between an input and output. At times, an accurate output is not determined by complete set of the input features instead, it is decided only by the subset of features. The probability of data mining algorithms extracting incorrect pattern is relatively high in these kind of multidimensional dataset. Most of data mining techniques need some degree of reduction in order to handle with large amounts of data, so an efficient and effective dimensional reduction method is required [7].

Feature selection is the choosing of a small subset of features that is necessary and sufficient to describe the target concept [8]. Selecting suitable set of features is extremely important since the feature set selected is the only source of information for any learning algorithm. The main aim of feature selection is to avoid selecting too many or too few features than is necessary. If too few features are selected, there is a good chance that the information content in this set of features is low. Instead, if too many (unrelated) features are selected, the effects due to noise present in the data may minimize the information content [9]. Two problems which may be induced by irrelevant features involved in machine learning process is the increased computational cost and also it leads to over fitting.

Feature selection is broadly classified into two types, namely the filter method and wrapper method. Filters estimate a relevance index for each feature to measure how relevant a feature is to the target. Then filter rank features by their relevance indices and perform search according to the ranks or based on some statistical criterion e.g. significance level. The most distinguishing characteristic of filters is that the relevance index is calculated based solely on a single feature without considering the values of other features.

Wrappers rank feature subsets by the prediction performance of a classifier on the given subset. Wrappers can be used to search through all possible subsets of features and explore the mutual information between features. After choosing a classifier (preferably consistent), wrappers evaluate the prediction performance either by cross-validation or theoretical performance bounds.

The feature selection procedure is addressed to select important feature and reduce the complexity of models. Feature selection is to determine a minimal feature subset of the problem domain while retaining high accuracy in representing the features. In this paper, ReliefF method is used to select important feature for wind speed forecasting.

#### 3.1 ReliefF method

The key idea ReliefF is to estimate the quality of attributes according to how well their values distinguish between instances that are near to each other. Given a randomly selected instance  $I_m$  from class  $L$ , ReliefF searches for  $k$  of its nearest neighbors from the same class called nearest hits  $H$ , and also  $k$  nearest neighbors from each of the different classes, called nearest misses  $M$ . It then updates the quality estimation  $W_i$  for  $i$ -th attribute based on their values for  $I_m$ ,  $H$ ,  $M$ . If instance  $I_m$  and those in  $H$  have different values on the  $i$ -th attribute, then the quality estimation  $W_i$  is decreased. On the other hand, if instance  $I_m$  and those in  $M$  have different values on the  $i$ -th attribute, then  $W_i$  is increased. These whole processes repeat  $n$  times to return weights of all features. ReliefF is fast, not limited to data types, fairly noise-tolerant and unaffected by feature interaction. The meteorological datasets contain following attributes like incoming shortwave radiation, reflected shortwave radiation, relative humidity, ambient air temperature, barometric pressure, wind direction, precipitation (Tipping Bucket), precipitation (Belfort), sonic range sensor for snow depth, shallow soil temperature, deep soil temperature, deep soil moisture, precipitation (GeoNor). The ReliefF feature selection algorithm is applied in this meteorological dataset for wind speed prediction. Features like ambient air temperature, wind direction, relative humidity, incoming and reflected shortwave are identified to be significant as by the algorithm and it is shown in Fig 2.

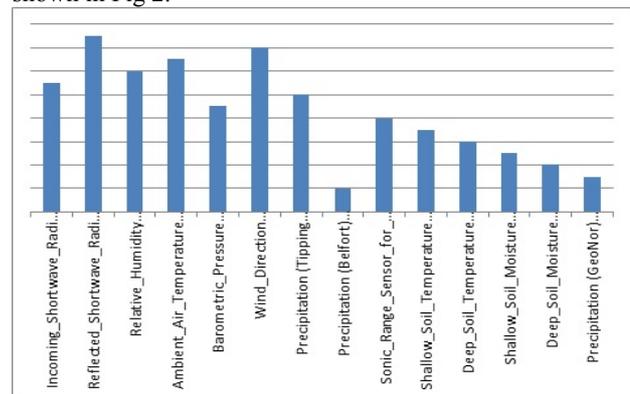


Fig. 2 Feature selection using ReliefF method

### 4. Artificial Neural Network (ANN)

ANN may be defined as an information processing model that is inspired by the way biological nervous system. The

main objective of the neural network research is to develop a computational device for modelling the brain to perform various computational tasks at a faster rate than the traditional systems. ANN have been successfully applied to a number of time series prediction and modelling tasks, including financial time series prediction [10], communication network traffic prediction [11], chaotic time series prediction [12],[13]. Neural networks have more comprehensive and more flexible functional forms compared to the traditional statistical methods. There are two types of neural networks: dynamic and static networks. [14],[15],[16].Static networks, such as Adaptive Neuro-Fuzzy Inference Systems and Multi-layer Feed-forward Neural Network, have no feedback, and the output are calculated directly based on their connection with Feed-forward input. But in dynamic neural networks, such as nonlinear neural network autoregressive (NNAR), the output depend on the current and past values of input, output, and the network structure [17].

**4.1 NARX model**

The nonlinear autoregressive network with exogenous input (NARX) is a recurrent dynamic network, with feedback connections enclosing several layers of the network. This model is well suited for modelling nonlinear systems and especially time series. The most important application of NARX dynamic neural networks is in control systems. There are two significant features of NARX. Learning is more effective in NARX networks than in other neural network (the gradient descent is better in NARX). Second, these networks converge much faster and generalize better than other networks [18], [19], [20].

NARX network can be trained in two modes: series parallel and only parallel. In series-parallel mode the true output is used instead of feeding back the estimated output. In parallel mode the output is feedback to the input of the feedforward neural network as part of the standard NARX architecture. [21], [22]. This paper focuses on parallel mode. The structure of NARX network used in the paper as shown in Fig.3.

The input for NARX network consists of two parts: the external input and the previous output of the network. We assume that  $w(t)$  and  $s(t)$  are the external input and output of network at time  $t$  respectively, noise delay time of output feedback  $n_1$  is the delay time of external input, thus the output of NARX network is defined as

$$s(t) = f(s(t - 1), s(t - 2) \dots s(t - n_0), w(t - 1), w(t - 2) \dots w(t - n_i)) \quad (1)$$

The output of network depends not only on  $w(t-i)$  where  $i=1, \dots, n_i$ , current and previous input to the network, but also on  $s(t-i)$ ,  $i=1, \dots, n_0$ , previous output of the network.[23],[24],[25].

**5. Results and Discussion**

The proposed NARX model is tested for the time series meteorological data set of 3000 samples having 15 minute interval from weather station at University of Waterloo. The proposed NARX model with all features (original datasets) is used to forecast the wind speed. ReliefF feature selection method is used to find the important features of wind speed

forecasting and reduce the complexity of model. NARX model with selected features like wind direction, temperature, humidity, pressure is used to forecast the wind speed.

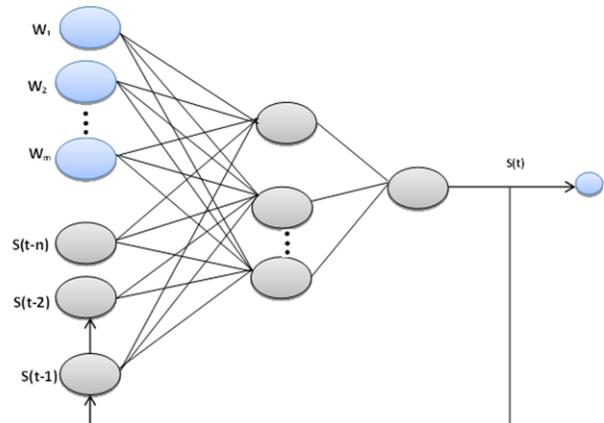


Fig. 3 NARX network structure

**5.1 Performance Evaluation**

The performance of the model in terms of the forecasting accuracy, several evaluation criteria is used. Mean Square Error (MSE) is a widely used criterion to find the difference between the values obtained by the proposed model ( $y_i$ ) and the actual values ( $t_i$ ),  $N$  is the number of forecasted samples. Lesser the value of MSE more is the accuracy of forecasting.

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i - y_i)^2 \quad (2)$$

The proposed NARX model with feature selection and without feature selection is tested. The wind speed forecast output of the NARX model without feature selection is shown in Fig 4. The wind speed forecast output of the NARX model with feature selection is shown in Fig 5.

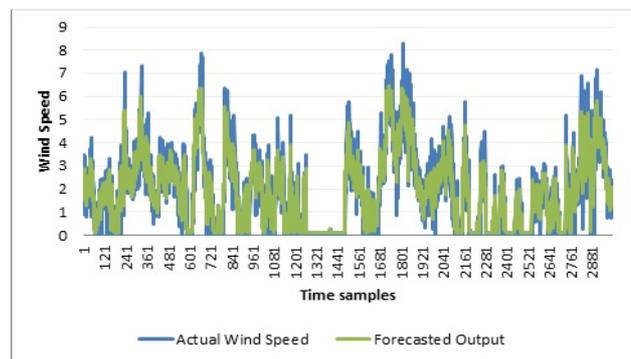


Fig 4. Actual and forecasted wind speed output using NARX model without feature selection

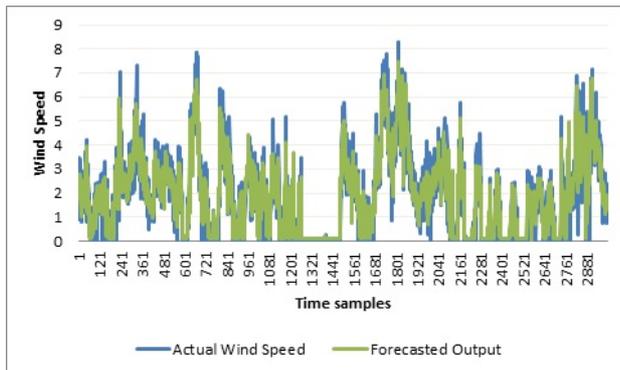


Fig 5. Actual and forecasted wind speed output using NARX model with feature selection.

It is clear from the result that the performance of NARX with feature selection mean square error is a 0.4771 and NARX model without feature selection MSE is 0.6355.

Forecasting wind speed using NARX model with feature selection is accurate.

## 6. Conclusion

In this paper, NARX model is implemented to forecast short term wind speed. The results indicate that the NARX with feature selection is the best fit model for the given data set. Wind speed has direct influence on the power generated by the wind turbine. By NARX model results the important feature selection for wind speed forecast are wind direction, temperature and humidity. Proper knowledge of wind speed and its forecast is used for scheduling and planning the wind power generation. We propose to implement data mining techniques as part of our future work to reduce the forecasting error.

## References

- Jaesung jung, Robert P.Broadwater, "Current status and future for wind speed and power forecasting", *Renewable and Sustainable Energy Reviews* 31(2014) pages 762-777.
- Monterio, R.Beesa, V.Miranda, A.Botterud,,J.Wang and G.Conzelmann, "Wind Power Forecasting: State-of-the Art 2009", Argonne National Laboratory, Nov(2009).
- Hui Liu, Chao Chen b, Hong-qi Tian , Yan-fei Li, "A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks", *Renewable Energy* 48 (2012) pages 545-556.
- K. GnanaSheela , S.N.Deepa, "Neural network based hybrid computing model for wind speed prediction", *Neurocomputing*, 122(2013) 425-429
- S. Salcedo-Sanz, A. Pastor-Sánchez , L. Prieto , A. Blanco-Aguilera , R. García-Herrera, "Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization – Extreme learning machine approach", *Energy Conversion and Management* 2014.
- Hui Liu , Hong-qi Tian , Yan-fei Li , "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction", *Applied Energy*, Applied Energy, 98 (2012) pages 415-424
- Richard Jensen, Qiang Shen, "Computational intelligence and feature selection", IEEE press series on computational intelligence, A John Wiley & Sons, Inc., Publication.(2008).
- Kira, K., Rendell, L.A., "A practical approach to feature selection", *Proceedings of the Ninth International Conference on Machine Learning*, pp. 249-256, (1992).
- Selwyn Piramuthu, "Evaluating feature selection methods for learning in data mining applications", *European Journal of Operational Research* 156 (2004) 483-494.
- Kim TY, Oh KJ, Kim C, Do JD, "Artificial neural networks for non-stationary time series", *Neurocomputing* (2004); 61:439-47.
- Bhattacharya, A.G. Parlos, A.F. Atiya, "Prediction of MPEG-coded video source traffic using recurrent neural networks", *IEEE Trans. Neural Networks* 51(8) (2003) pages 2177-2190.
- R.J. Duro, J.S. Reyes, "Discrete-time backpropagation for training synaptic delay-based artificial neural networks", *IEEE Transactions on neural networks*; 10(4): 779-789, (1999).
- Hang xie, Hao tang, Yu-he liao, "Time series prediction based on narx neural networks: an advanced approach", *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding*, 12-15 July (2009).
- Abbas Ali Abounoori ,Hanieh Mohammadali , Nadiya Gandali Alikhani and Esmail Naderi, "Comparative study of static and dynamic neural network models for nonlinear time series forecasting", *Munich Personal RePEc Archive*, (2012).
- Deng,J "Dynamic Neural Networks with Hybrid Structures for Nonlinear System Identification", *Engineering Applications of Artificial Intelligence* 26 (1), PP. 281-292, (2013).
- Chiang, Y. M., Chang, L. Ch., Chang, F.J. , "Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modelling", *Journal of Hydrology*, 290(3-4), PP. 297-311, (2004)
- Tsoi, A. Ch., Back, A., "Static and dynamic preprocessing methods in neural networks", *Engineering Applications of Artificial Intelligence* 8(6), PP. 633-642, (1995).
- Eugen Diaconescu, "The use of NARX Neural Networks to predict Chaotic Time Series", *WSEAS Transactions on Computer Research*, Issue 3, Volume 3, March(2008).
- Simon Haykin, "Neural Networks", Second Edition, Pearson Education, (1999).
- Tsungnan Lin, Bill G. Horne, Peter Tino, C. Lee Giles, "Learning long-term dependencies in NARX recurrent neural networks", *IEEE Transactions on Neural Networks*, Vol. 7, No. 6,(1996) pp. 1329-1351.
- Yang Gao, Meng Joo Er, "NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches", *Fuzzy Sets and Systems*, Vol. 150, No. 2, 2005, pp.331-350.
- Hang xie, Hao tang, Yu-he liao, "Time series prediction based on narx neural networks: an advanced approach", *proceedings of the eighth international conference on machine learning and cybernetics, baoding*, 12-15 July (2009).
- [www.mathworks.com/help/nnet/ug/design-time-series-narx-feedback-neural-networks.html](http://www.mathworks.com/help/nnet/ug/design-time-series-narx-feedback-neural-networks.html)
- M. Ramasamy, H. Zabiri, N.T.Thanh ha, N.M. Ramli, "Heat Exchanger Performance Prediction Modeling using NARX-type Neural Networks", *Proceedings of the WSEAS Int. Conf. on Waste Management, Water Pollution, Air Pollution, Indoor Climate, Arcachon, France*, October 14-16, (2007).
- Dr. Rafid Ahmed Khalil, "Comparison of Four Neural Network Learning Methods Based on Genetic Algorithm for Non-linear Dynamic Systems Identification", *AL Rafdain Engineering Journal*, Vol.20, Issue1, (2012).