Forecasting of Wind Speed Using Feature Selection and Neural Networks

Senthil Kumar P*‡, Daphne Lopez**,

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

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

(senbe@rediffmail.com, daphnelopez@vit.ac.in)

‡Corresponding Author;Senthil Kumar P, VIT University, Vellore, Tamil Nadu, India, Tel: +91 9994626135

senbe@rediffmail.com

Received: 15.03.2016 Accepted:10.06.2016

Abstract-Wind energy is rapidly increasing and it is becoming a significant contributor to the electricity grid. Wind speed is an important factor in wind power production and integration. This paper presents a wind speed forecasting using feature selection method and bagging neural network. Feature selection plays an essential role in the machine learning environment and especially in the prediction task. The ReliefF feature selection method is used for identification of necessary features for wind speed forecast and reduces the complexity of the model. A detailed investigation is to forecast wind speed with meteorological time series data as input variable using a bagging neural network. Performance is evaluated in terms of mean square error when using the feature selection method with the bagging neural network.

Keyword: Bagging, Feature selection, Forecasting, Neural network, ReliefF, Wind speed.

1. Introduction

Wind energy has become an important source of alternative energy in recent years. It has more advantages with respect to other sources in terms of installation and generation cost. The wind speed forecast in the wind energy sector is important for the following reasons like wind power system planning for unit commitment decision, load balancing decision, maintenance arrangement and energy storage capacity optimization. Wind speed is periodarrangementinformation measured at various interims of time.

In general wind power sector, forecasting can be divided into four time horizon, as very short, short,medium and long term. Very short term time horizon is for very few minutes only and these forecasts are useful for turbine control and unit commitment. Short term time horizon could be for 30 minutes to 6 hours. Theseforecasts are useful for preload sharing. Medium term time horizon is for a 6 hours to 24 hours. These forecasts are useful for power system management and energy trading. Long term time horizon is for one day to seven days. These forecasts are useful for maintenance planningof conventional power generation plants.[1], [2], [3].

Forecasting becomes an important factor when penetration of wind speed increases. Accurate forecasting reduces the risk of uncertainty while integrating power system. Wind speed forecasting provides details about the amount of wind power generated in a particular interval of time. Mohammed monafered et al., [4] proposed wind speed forecasting utilizing fuzzy logic and artificial neural network. This strategy gives a fuzzy rule base for fuzzylogic and quick learning process for neural systems. This model indicated less computational time and better wind speed prediction compared with other traditional model.Jujie Wang et al., [5] have proposed a hybrid model forwind speed prediction. This model is based on the seasonal adjustment, exponential smoothing and radial basis function neural network. The model numerically outperformed compare to other models. The proposed methodology was effective in enhancing the forecast exactness. XiaobingKonga et al. [6] propose a wind speed prediction using Reduced Support Vector Machine (RSVM). The principal component analysis method is to diminish the measurement of the observed variables and improve the model's generalization power in wind speed RSVM modelling.The RSVM parameters are improved byutilizing a particle swarm optimization algorithm. The proposed RSVM result is promising for wind speed forecasting.

2. Methodology

2.1.Feature Selection

The feature selection method chooses a subset of variables from the inputs which can proficiently portray the input data while diminishing the impact from the irrelevant variable and provide good forecast results. It helps in comprehension information , decreasing computation requirement, lowering the curse of dimensionality and improving the predictor overall performance. [7], [8].

Feature selection increases the accuracy level and reduces data training time. From the machine learning point of view, the removal of the redundant and irrelevant attributes from the dataset before evaluating algorithm is important. Keeping irrelevant attributes in the dataset can result in overfitting. This overfitting of the training data can affect the prediction accuracy. Feature selection is not comparing with other dimensionality reduction method such as principle component analysis. Feature selection and dimensionality reduction methods reduce the number of attributes in the dataset, but dimensional reduction creates a new combination of attributes, whereas feature selection method reduce the attributes present in the dataset without changing them.[9], [10], [11]. The general procedure for feature selection process as shown in fig. 1.

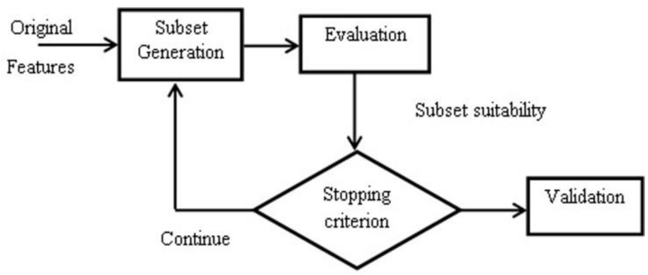

Fig. 1. Feature selection process

There are three grouping of feature selection methods like the filter, the wrapper and the hybrid methods. Filter methods, practice a statistical degree to allocate a ranking to every feature. Ranking measure is applied to attain the variables and a threshold is used to evacuate variables underneath the edge. Filter methods have a tendency to choose subsets with all the capabilities and therefore a right threshold is needed to choose a subset. Examples of few filter techniques include the chi square test, mutual information gain, correlation coefficient, Relief method, focus method, Entropy Based Reduction (EBR), Fractal Dimension Reduction (FDR). A Wrapper method is the selection of a set of features, where distinctive mixes are orchestrated, assessed and contrasted with different blends. A predictive model is utilized for assessment of a grouping of features and assigns a score based on model exactness. In the wrapper method, a search strategy iteratively adds or eliminates features from the data to search a most ideal feature subset that improves accuracy.Search strategies are used by the wrapper method are forward selection and forward elimination. Examples of the wrapper method include the genetic algorithm, Lag Vegas Wrapper (LVW). Embedded methods are more suitable for feature selection mainly on high dimensional data. These strategies incorporate both the intrinsic characteristic of data sets and the predefined mining algorithm. [12]. In this paper, ReliefF method is used to select important feature for wind speed forecasting. ReliefF is to approximate the quality of attributes as indicated by the way in which their values differentiate between occurrences that are near each other. ReliefF randomly chose an instances from class R_i, but then searches for k of its from the same class, called nearest hits and k-nearest neighbours from each of the diverse classes, called nearest misses. It updates the quality estimation W_i for all attributes relying on their values for R_i, hits and misses. If instance R_i and those in hits have diverse values on the i -th attribute, then the quality estimation W_i is lesser. Alternatively, when instance R_i and those in misses have diverse values on the i-th attribute, W_i is higher. This entire process is rehashed n times to return weights of all features. ReliefF is quick, not constrained to data types, reasonably noise-tolerant and unaffected by feature interaction. [13], [14], [22], [23].

2.2. Artificial Neural Networks (ANN)

ANN may be described as an information processing model that is inspired by means of the biological nervous system. The fundamental motivation behind the neural system examination is to progress a computational device for modelling the brain to carry out numerous computational tasks at a faster rate than the conventional systems. ANN performs various tasks such as classification, prediction, optimization, identification, evaluation and control of complex systems. The primary model of neural network includes a large number of interrelated processing elements (nodes). Some of the existing types of neuron interconnection architecture are single or multilayer feed-forward network, single node with its own feedback, single-layer recurrent network and multilayer recurrent network. [15], [16], A sample of multilayer feed-forward network as shown in figure 2.

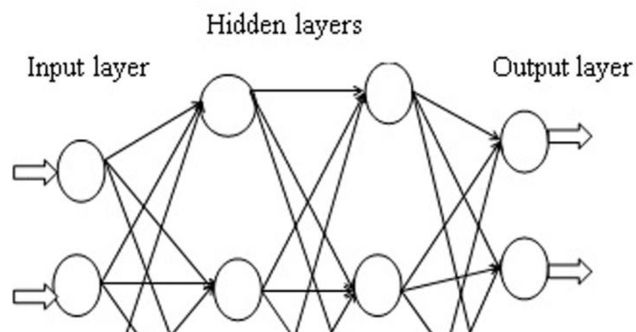

Fig. 2. Multilayer feed-forward networks

These processing elements of the ANN have ability to learn and generalize from the given data by adjustment of weights and make useful decisions. The procedure begins with a summation of weighted activation of different neurons its approaching associations. At that point the weighted aggregate is passed through an activation function and this activated value is the output of the neuron. There are several activation functions like to identify the function, binary step,bipolar step, sigmoid and ramp function. The sigmoid functions are widely used as activation function.

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

The framework of the proposed wind speed forecasting architecture is shown in figure.3

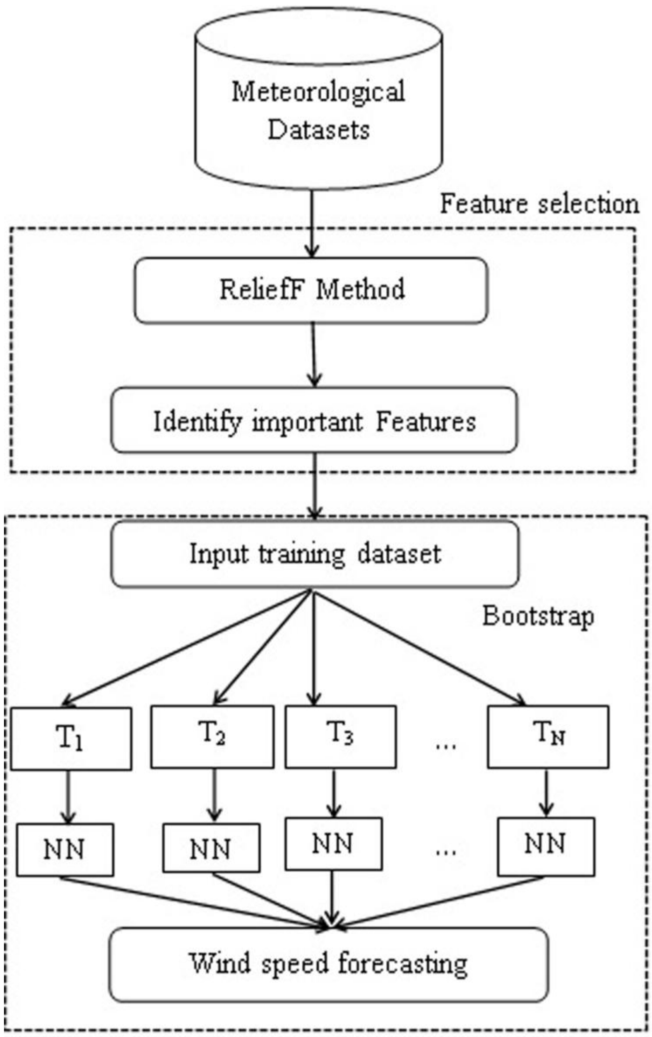

Fig. 3. Framework of wind speed forecasting

2.3. Wind speed forecasting using Bagging Neural Network

2.3.1. Ensemble methods

An ensemble method is widely used in the area of machine learning and statistics. Ensemble methods train multiple learners to solve the same problem. In ensemblemethods numbers of learners share the input and combine to produce an overall output. Figure 4. shows the structure of ensemble method which generally improves the generalization performance of a set of classifier on a domain. The combiner output of the number of classifier reduces the risk of selecting a weak performance classifier. [17].

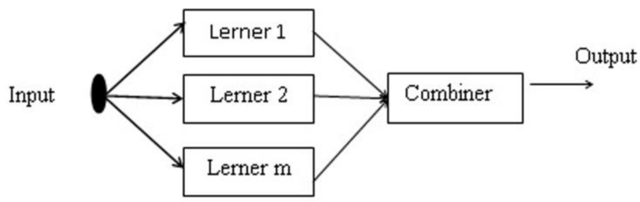

Fig. 4. Structure of classifier ensemble

Handling a large volume of data by a single classifier is rather difficult. So training of different classifiers on different partitions of data is necessary. Applications of ensemble methods in various fields like computer vision, computer security problems, intrusion detection and computer aided medical diagnosis. Bagging and boosting are widely used in machine learning to improve the performance of prediction and classification algorithms [18].

The name Bagging is derived from the Bootstrap aggregation. It is an ensemble method of combining multiple predictors. Bagging algorithm is used to exploit the independence between the learners and also reduce error by combining independent learners. Bagging algorithm is mainly designed for increasing the performance of machine learning algorithms. Bagging works admirably for unstable learning algorithms like neural systems, decision tree and regression while bagging can slightly reduce performance of stable learning algorithms like K- nearest neighbor. Bagging algorithm adopts the bootstrap distribution for generating different base learners. Given a training dataset contains an N number of training examples and m number of samples are generated. It is possible that some of the original samples are repeated more than once in the training samples while some of the original examples are not present in the training example. The different training samples are used to train and make different models. Finally the bagging combines the output obtained from the various models. [19],[20],[21].

Consider the training data are denoted by $N^t_{training=\left(x^t_{training},y^t_{training}\right)where t=1.....N}$. $x^t_{training}$ is an N-dimensional features for the i[th] instance and $y^t_{training}$ is the wind speed forecast for the i[th] time instance. $x^t_{training}$ consist of the following features like wind direction, temperature, humidity, etc., at the i[th] time instance. The test set is given as $N^t_{test=(x^t_{test},y^t_{test})where t=1.....N'}$ Given the test features X_{test}, predictor for P[y_{test}]= $f(X_{test})$

The steps involved in bagging neural network algorithm are summarized as under:

1. Fix the total number of iteration equal to the total number of different ANN and the number of bootstrap samples.
2. Construct at each iteration, we construct a bootstrap sampling for training the base learners. Given a training dataset contains N number of training examples and m number of samples will be generated by sampling with replacement.
3. Compute the bootstrap predictor by using ANN.
4. At the end of the iterations, the bagged predictor is obtained as

$$f_B(X_{test}) = P[f_{1*}(X_{test}), f_{2*}(X_{test}) \ldots . f_{n*}(X_{test})] \quad (2)$$

Calculate $f_B(X_{test}) \approx \frac{1}{D}\sum_{n=1}^{D} f_{n*}(X_{test})$ where D is number of neural network.

5. The forecasted value is
$$y_{test} = f_B(X_{test}) \quad (3)$$

Bagging is aggregation of misclassification errors on different data splits which gives a reliable estimate of the predictive ability of a learning method.

3. Results and Discussion

The proposed model is tested with the time series meteorological dataset of 3000 samples with a 15 minute interval is collected from the weather station at the University of Waterloo. The meteorological datasets contain attributes like wind speed (actual wind speed), 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 soiltempearture, deep soil moisture, precipitation. Fig. 5 shows the actual wind speed data for 3000 samples at intervals of 15 minutes.

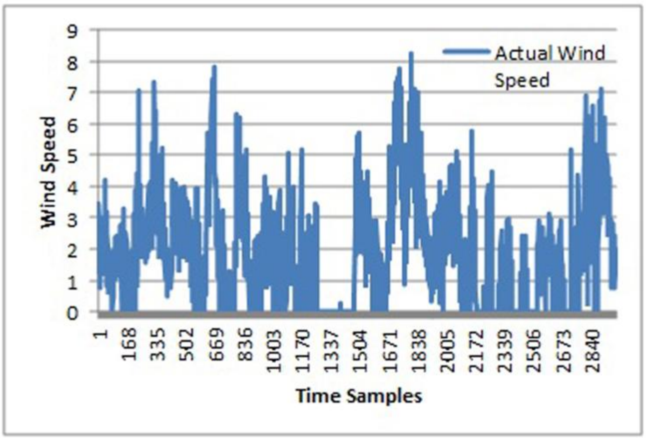

Fig. 5. Actual wind speed data

The ReliefF feature selection algorithm is applied in this meteorological dataset for wind speed prediction. Features like ambient air temperature, wind direction, relative humidity, pressure, incoming and reflected shortwave are identified to be significant as by the algorithm. ReliefF feature selection method is used for finding the important features of wind speed forecasting and reduce the complexity of the model. The baggedneural network model with selected features like wind direction, temperature, humidity, pressure, incoming and reflected shortwave is used to predict the wind speed. A significant set of feature as input is given. These features consist of meteorological weather data for wind speed prediction over a time period. We also have a known wind speed output corresponding to the wind speed over a time period.

3.1.Performance evaluation

Several evaluation criteria are used for study of the performance of the model in terms of forecast accuracy. Mean Square Error (MSE) is a broadly utilized criterion to find the difference among the values acquired by way of the proposed model (yi) and the actual values (ti), N is the number of forecasted samples. Lesser the value of MSE more is the accuracy of forecasting.

$$\text{MSE} = \frac{1}{N}\sum_{1=1}^{N}(t_i - y_i)^2 \quad (4)$$

The proposed bagged neural network model with feature selection and without feature selection is tested. The wind speed forecast output of the bagged neural network without feature selection as shown in Figure 6. The wind speed forecast output of the bagged neural network with feature selection as shown in Figure 7.

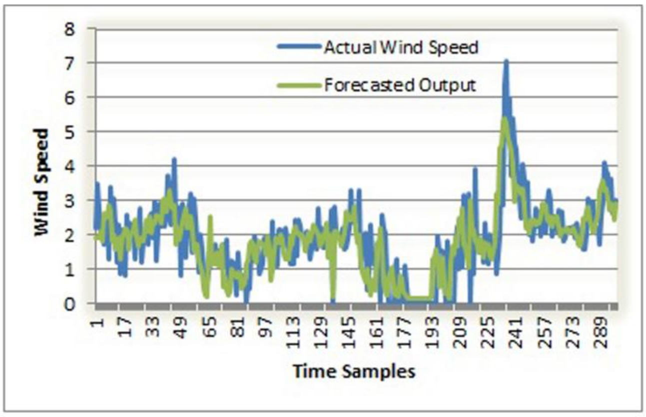

Fig. 6. Actual and forecasted wind speed output using bagged neural network model without feature selection.

The result clearly shows that the performance of the bagged neural network model with feature selection mean square error is a 0.4331 and the bagged neural network model without feature selection MSE is 0.6345. Forecasting wind speed using a bagged neural network with feature selection method is more accurate.

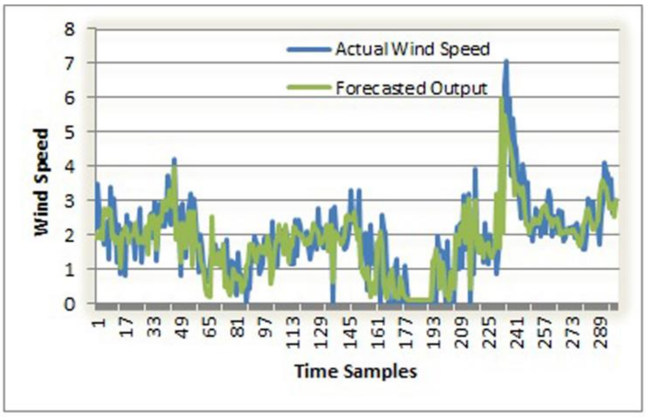

Fig. 7. Actual and forecasted wind speed output using bagged neural network with feature selection

4. Conclusion

In this paper, we first present ReliefF feature selection method is used for identification of significant features for wind speed forecast and reduce the complexity of the model. By ReliefF results the important features for wind speed forecast are wind direction, temperature and humidity etc,. Next step, the bagged neural network is implemented to forecast wind speed. The bagged neural network can be used to improve wind speed forecasting accuracy significantly. The results indicate the bagged neural network with feature selection as the best fit model for the given dataset. Wind speed has a direct influence on the power generated by the wind turbine. A proper knowledge of wind speed and its forecast is used for scheduling and planning the wind power generation. We propose to implement other data mining techniques as part of our future work to reduce the forecasting error.

References

[1] Jaesungjung, Robert P.Broadwater, “Current status and future for wind speed and power forecasting”, Renewable and Sustainable Energy Reviews Vol .31,pp. 762-777, 2014.

[2] Monterio, R.Beesa, V.Miranda, A.Botterud,,J.Wang and G.Conzelmarn, “Wind Power Forecasting: State-of-the Art 2009”, Argonne National Laboratory, Nov 2009.

[3] MelamBhaskar, Amit Jain, N. Venkatasrinath, “Wind Speed Forecasting: Present Status”, International Conference on Power Technology, 2010.

[4] Mohammad Monfared, Hasan Rastegar, Hossein MadadiKojabadi,”A new strategy for wind speed forecasting using artificial intelligent methods”, Renewable Energy, Vol.34, pp.845–848, 2009.

[5] Jujie Wang , Wenyu Zhang, Jianzhou Wan, Tingting Han, Lingbin Kong,” A novel hybrid approach for wind speed prediction”, Information Sciences Vol. 273, pp. 304–318, 2014.

[6] Xiaobing Kong, XiangjieLiua,n, RuifengShi, KwangY.Lee, “Wind speed prediction using reduced support vector machines with feature selection”, Neurocomputing, Vol.169,pp. 449–456, 2015.

[7] Richard Jensen, Qiang Shen, “ Computational intelligence and feature selection”, IEEE press series on computational intelligence, A John Wiley & Sons, Inc., Publication.(2008).

[8] Kira, K., Rendell, L.A., “A practical approach to feature selection”, Proceedings of the Ninth International Conference on Machine Learning. pp. 249–256, (1992).

[9] Selwyn Piramuthu, “Evaluating feature selection methods for learning in data mining applications”, European Journal of Operational Research 156 (2004) 483–494.

[10] Girish Chandrashekar and FeratSahin, A survey on feature selection methods, computer and Electrical Engineering Vol.40, pp. 16-28,2014.

[11] Vipinkumar and sonajhariaminz, feature selection: A literature review, Smart computing Review, vol 4, No 3, pp.211-229, 2014.

[12] Pramod Kumar Singh , Dimension Reduction Methods in High Dimensional Data Mining, Available: www.sau.ac.in/~mlta/studymaterial/22-12-2013/PKSingh.pdf

[13] Marko Robnik-ˇSikonja and Igor Kononenko , “Theoretical and Empirical Analysis of ReliefF and RReliefF”, Machine Learning Journal Vol. 53, pp.:23-69, 2003.

[14] Liuzhi Yin, Yong Ge, Keli Xiao, Xuehua Wang , XiaojunQuan , Feature selection for high-dimensional imbalanced data, Vol.105, pp3-11, 2013.

[15] S.N. Sivanandam and S.N. Deepa, “Principle of Soft computing”, Wiley India (P)Ltd,2008.

[16] Rasit Ata, “Artificial neural networks applications in wind energy systems: a review”, Renewable and Sustainable Energy Reviews , Vol.49,pp.534–562, 2015

[17] Zhi-Hua Zhou, “Ensemble Methods Foundations and Algorithms”, CRC Press Taylor & Francis Group 2012.

[18] Chih-Fong Tsai,Yu-FengHsu,DavidC.Yen, A comparative study of classfier ensembles for bankruptcy prediction, Vol.24,pp977-984, 2014.

[19] A.S. Khwaja, M. Naeem, A. Anpalagan, A. Venetsanopoulos, B. Venkatesh, “Improved short-term load forecasting using bagged neural networks”, Electric Power Systems Research, Vol. 125,pp. 109–115, 2015.

[20] E. Shafiei1, H. Jazayeri-Rad, Improving the Identification Performance of an Industrial Process Using Multiple Neural Networks, American Journal of Intelligent Systems, Vol. 2, Issue 4, pp. 40-44, 2012.

[21] HuanLong,ZZijun Zhang Yan Su, Analysis of daily solar power prediction with data-driven approaches, Applied Energy, Vol.126,pp29-37, 2014.

[22] Reliefalgorithm,http://users.cis.fiu.edu/~yzhan004/Algorithm/reliefF.html.

[23] Marko Robnik-Sikonja, Igor Kononenko, Theoretical and Empirical Analysis of ReliefF and RReliefF, Machine Learning Journal (2003)53:23-69.