Volume 6, Issue 12, December 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Feature Selection in Artificial Neural Network Model on NOx Prediction from Utility Boilers

K. S. Madhavan*
Sr. DGM / PCS, Corp. R&D, BHEL,
India

Abstract— Recent trends demonstrate that as boiler technologies grow more sophisticated, their control technologies must keep pace to take advantage of increasing capabilities. Today's supercritical units can achieve thermal efficiency of up to 46% compared to a sub-critical plant's 30-38%. Increased efficiencies translate into reduced fuel costs and emissions. In this paper an attempt is made to identify all the parameters affecting NOx emissions from utility boilers and ranking them based on their sensitivities in an Artificial Neural Network model called the Multi-Layer Perceptron. The NOx emission prediction is done based on this ranking. Dominant parameters (features) may be identified and the Artificial Neural Network may be retrained to get better values of prediction. The quality of prediction is represented by value of the correlation coefficient.

Keywords— NOx, Emissions, Prediction, Sensitivity, Ranking, Artificial Neural Network, Multi-Layer Perceptron, Correlation Coefficient, Genetic Algorithm.

I. INTRODUCTION

Nitrogen oxides consist mostly of NO with much lower levels of NO_2 and N_2O [1]. NO is formed as thermal NO, prompt NO, and fuel NO. Earliest research on NOx formation can be traced to Zeldovich who first postulated the thermal NO formation mechanism. NOx emission reduction can be primarily combustion based or alternately due to flue gas treatment. The fundamental equations for NOx formation cannot be directly applied as many of the variables cannot be measured online or reliably measured. No model structure is available. Information such as steady state relationships between input and output is not readily available. This causes a problem in the lumped system model or Wiener-Hammerstein model of the process.

Fuel NO contributed over 75% of the total emissions under all possible conditions [2]. Fuel NO formation is a complex process. During de-volatilisation, fuel-N is split into char-N and volatile-N that mainly includes tar-N, HCN and NH3. The distribution of N depends on the type of coal and temperature. The final NO production decreases with increasing bulk gas NO concentration, char particle size, char loading, and char reactivity.

Even with all this knowledge, NOx emission phenomena is understood only to a limited extent. More and more studies have to be conducted on the subject and model understanding has to improve. NOx measurement is presently done through Differential Optical Absorption Spectroscopy (DOAS) techniques and reduction in NOx levels may be brought about through Optimization of mills, low NOx burners, staged combustion, biasing the fuel and changing the Excess Air settings. In this paper we are attempting to generate a feature selection process for the Artificial Neural Network (ANN) prediction model on NOx emissions from utility boilers.

Around 96 field data consisting of NOx and few required influencing parameters were taken and analyzed for the sensitivity importance of each influencing factor following weights method and first order derivatives method of ANN and the results are discussed in this paper.

Due to lack of sound model understanding of the NOx emission prediction phenomena, we have to first identify all or few dominant features that affect NOx emission. As a result of this exercise we have identified, Furnace Width in mm, Furnace Depth in mm, Height from topmost mill fired to bottom of platen superheater in mm, Load, Total Coal Flow, Primary Air, Secondary Air, Burner Tilt, Mill Combination, Excess Air, results of proximate and ultimate analysis of fuel, coal gross calorific value (gcv). Mill combination and no. of mills fired were combined into a single input by arranging the mill combination in 10 bit pattern and setting the bit to 1 if mill is fired otherwise setting the bit to 0 if mill is not fired. NOx values have been measured at two points: One on the stack and one in the economizer outlet to facilitate implementation of deNOx systems.

Once feature selection was broadly done based on domain expertise, an exercise to identify feature sensitivity or feature ranking was taken up. In this process the dominant features affecting NOx emission were also identified. Features gathered from the observation of a phenomenon are not all equally informative: some of them may be noisy, correlated or irrelevant. Feature selection aims at extracting a feature set that is relevant for a given task. This problem is complex and remains an important issue in many domains. In the field of neural networks, feature selection has been studied for the last ten years and classical as well as original methods have been employed.

II. METHODOLOGY FOR FEATURE RANKING USING ARTIFICIAL NEURAL NETWORK

Feature Ranking in Artificial Neural Networks may be done using weight pruning, the first order derivative of the output, the second order derivative of the output or genetic algorithm method. The first three mentioned are based on gradient methods while the latter is based on non-gradient method. Second order methods use the Hessian Matrix or the second derivative of the outputs to compute sensitivity of the feature. These methods are more cumbersome, take more time to calculate, involve more computational complexity and are not used in this paper. The genetic algorithm [3] uses a robust statistical estimator to find the relative importance of inputs.

A. Weights Method

Weight Pruning can be done using the product of the weights from input layer to the hidden layer and from the hidden layer to the output layer, normalized by the entire sum of weight magnitudes. The ranking due to weight pruning is indicated by (1), [5].

$$S_i = \sum_{j \in H} \left(\frac{|w_{ji}|}{\sum_{i' \in I} |w_{ji'}|} \sum_{k \in O} \frac{|w_{kj}|}{\sum_{j' \in H} |w_{kj'}|} \right) \qquad \text{……..(1)}$$

w_{ji} = Weights from hidden layer to input layer
w_{kj} = Weights from output layer to hidden layer

where I, H, O denote respectively the input, hidden and output layer. Denominators in (1) which are simple normalizing factors will reduce to a value of 1 for linear activation functions. This has been vividly illustrated [6] in Figure 1. In such cases the entire (1) will reduce to a simplified form illustrated by (2).

For a better understanding of this measure, let us suppose that each hidden and output unit incoming weight vector has a unitary L1 norm, the above equation can be written as [4]:

$$S_i = \sum_{o \in O} \sum_{j \in H} |w_{oj} w_{ji}| \qquad \text{……..…(2)}$$

In (2), the inner term is the product of the weights from input i to hidden unit j and from j to output o. The importance of variable i for output o is the sum of the absolute values of these products over all the paths in the ANN from unit i to unit o. The importance of variable i is then defined as the sum of these values over all the outputs.

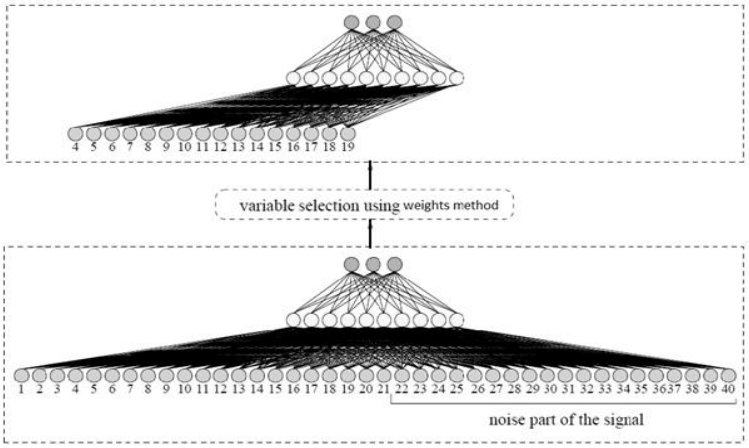

Fig. 1: Illustrating weight pruning for feature selection

After the weights are captured at the end of the run, (2) was applied by using NeurOn-line program to evaluate the product of the transpose of the weight matrices from input layer to hidden layer and from the hidden layer to the output layer. The feature sensitivity thus obtained is shown in Figure 2.

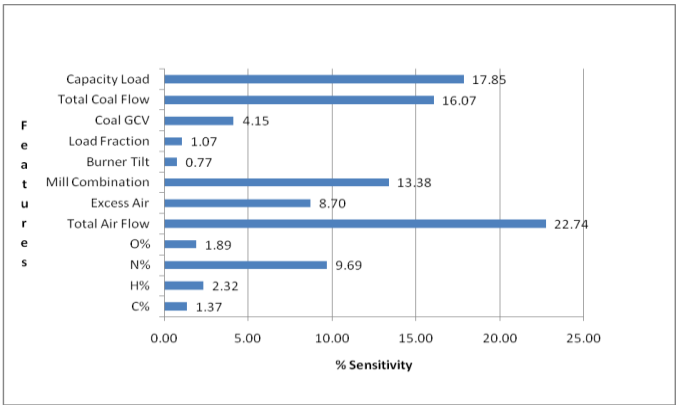

Fig. 2: Percentage Sensitivity showing feature ranking using simplified weights method

The graph showing predicted value of NOx versus actual value by using Weights Method of calculating sensitivity is shown in Figure 3. Correlation Coefficient is of the order of 0.92.

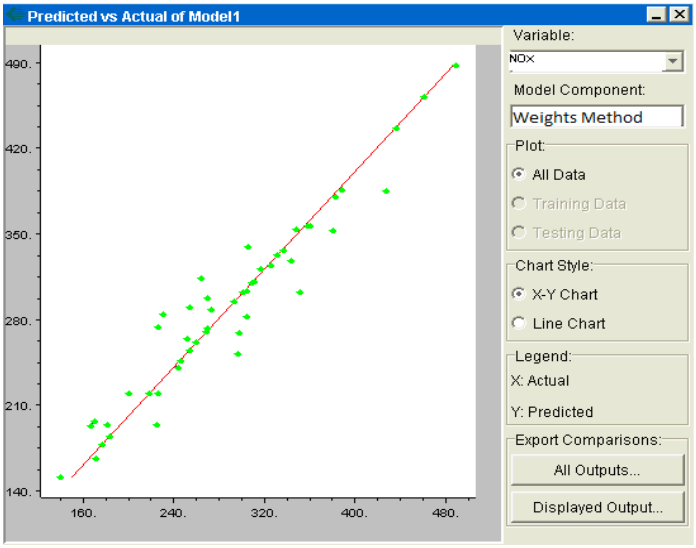

Fig. 3: Predicted Vs. Actual (Weights Method)

B. First Order Derivative Method

A second order BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm is being used to train the ANN in this approach.

For feature selection a first order derivative method is used. By progressively calculating the summation first order derivatives of the activation function at each stage, taking the difference between them in proportion of the square of the error function and normalising them based on the standard deviation of the first order derivatives, a measure of feature sensitivity is evolved [4].

In this paper the method used to compute sensitivity / ranking based on output derivatives is a normalized standard deviation of the derivatives. The relevant governing equation is shown in (3) [6].

$$S_i = \frac{1}{N^{1/2}} \frac{\left(\sum_l \left(\frac{\partial f}{\partial x_i}(x^l) - \sum_j \frac{\partial f}{\partial x_i}(x^j) \right)^2 \right)^{1/2}}{\sum_l \frac{\partial f}{\partial x_i}(x^l)} \qquad \text{……..(3)}$$

The feature sensitivity obtained by first order method using Multi-Layer Perceptron (MLP) is illustrated in the 2D bar chart shown in Figure 5. The values are shown at the right end of the bar chart. They indicate percentage contribution of each of the individual features fed into the Artificial Neural Network model using MLP. The percentage value shows the ranking of individual features in the MLP model. The graph showing predicted value of NOx versus actual value by using first order derivatives method of calculating sensitivity is shown in Figure 4. Dominant features are identified [7]. Correlation Coefficient which is the measure of the match between the predicted and actual distributions is of the order of 0.99.

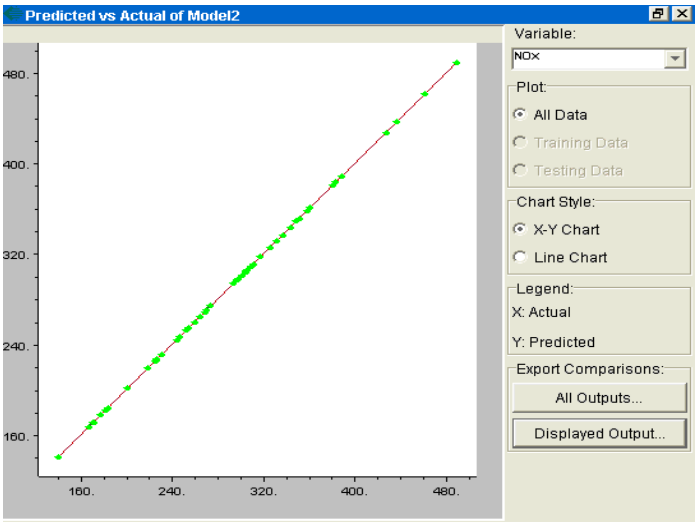

Fig. 4: Predicted Vs. Actual (First Order Derivative Method)

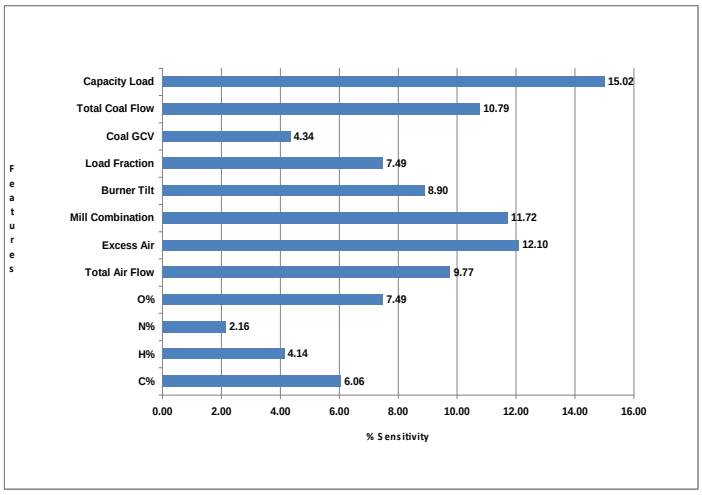

Fig. 5: Percentage Sensitivity showing feature ranking using first order derivative method

III. INFERENCES BASED ON MODEL UNDERSTANDING

The first order partial derivative method was found to be very useful as it gave the most complete results. The Weights method allowed a good classification of the input parameters.

IV. FUTURE WORK

The process of feature selection and prediction through Artificial Neural Network technique may be made online to capture dynamically the NOx variations and implement NOx emission reduction methods effectively in the Continuous Emission Monitoring Systems.

V. CONCLUSIONS

Feature Sensitivity of Multi-Layer Perceptron model for NOx emission prediction has been identified using Weights Method and First Order Derivative Method. The results have been compared and First Order Derivative Method has been found to be more consistent and accurate during the predictions. Dominant features are identified.

ACKNOWLEDGMENT

The authors are thankful to the BHEL management for rendering adequate support and facilities for carrying out the research published.

REFERENCES

[1] Zhangfa Wu, Fundamentals of pulverized coal combustion, CCC/95, ISBN 92-9029-410-8, 36 pp, March 2005.
[2] Hyung-Taek Kim and Wongee Chun, The particle size effect on the pollutant formation in pulverized coal combustion, Environmental Engineering Research, Vol. 3, No. 1,pp 21-29, 1998.
[3] GeneHunter (2015) online manual by Ward Systems Group, Maryland, USA.
[4] Hashem, S. Sensitivity Analysis for Feedforward Artificial Neural Networks with Differentiable Activation Function, Proceedings of 1992 International Joint Conference on Neural Networks IJCNN92 pp 419-424, 1992.
[5] Philippe Leray and Patrick Gallinari, Feature Selection with Neural Networks, Behaviormetrika, vol.26, pp 16-42, 1998.
[6] Yacoub, M and Bennani, Y, Features selection and architecture optimization in connectionist systems, International Journal of Neural Systems, Vol. 10, No. 5, 379-395, 2000.
[7] Devaraj, J. Preetha Roselyn, R. Uma Rani, Artificial neural network model for voltage security based contingency ranking, ScienceDirect, Applied Soft Computing pp 722–727, 2007.

