International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(1): 43-47

https://www.mathsjournal.com
Received: 16-11-2022
Accepted: 21-01-2023

Maibam Sanju Meetei
Assistant Professor,
Department of Electronics and Communication Engineering,
Rajiv Gandhi University,
Doimukh, Arunachal Pradesh,
India

Corresponding Author:
Maibam Sanju Meetei
Assistant Professor,
Department of Electronics and Communication Engineering,
Rajiv Gandhi University,
Doimukh, Arunachal Pradesh,
India

A comparative study of different multi-layer preceptorn learning algorithm with raw and normalized for classification of methane and hydrogen

Maibam Sanju Meetei

Abstract

In this study, the classification of 0.5% methane gas, 1% methane gas, 0.5% hydrogen gas, and 1% hydrogen gas is done using the multi-layer perceptron neural network. The neural network with one hidden layer is sufficient to classify the data in theoretically but it is number of neurons, training period is high and low convergence rate. So a neural network with two hidden layer is consider for the study as this network has higher convergence rate with lesser number of neurons. Out of eleven different learning algorithms, Levenberg-Marquardt learning algorithm shows the best algorithm for this study. Additionally, the impact of training with raw data versus training with normalized data is examined, and it is found that training with normalized data performs better because the neural network architecture has less number of neuron and higher performance i.e. mean square error is smaller. This research clearly demonstrates that MLP may be used to classify methane and hydrogen at various concentration levels. For raw data training, the numbers of neurons are fourteen and twelve neurons in first and second hidden layers respectively with a minimum MSE of 0.00586. For normalized data training the numbers of neurons are ten and eleven neurons in first and second hidden layers respectively with a minimum MSE of 0.002.

Keywords: Activation function, hidden neurons, learning algorithm, normalization, MSE

1. Introduction

The majority of gas recognition systems are PC-based and therefore not portable. To completely overcome the shortcomings of the cumbersome PC-based gas pattern recognizer, an embedded system that automatically reads the sensor input from the array, normalizes it, and computes the neural network output for the set of sensor values can be developed.

Several methods for pattern classification have been employed by many scientists and researchers, including principle component analysis (PCA) [1], partial least squares analysis (PLS) [2-3], and artificial neural networks (ANN)[4] using metal-oxide semiconductor (MOS) gas sensor arrays. Artificial neural networks, however, are said to perform better than existing recognition methods when quantitative or multi-component mixture classifications must be done [5]. The two most popular neural network techniques for pattern recognition are multilayer perceptron (MLP) and radial basis function (RBF). Moreover, it has been claimed that MLP has performed better than RBF [5-6].

Contrarily, larger networks can provide accurate approximations and the desired accuracy, but an overly complex structure may result in overfitting the training data, which performs well for patterns found in the training set but very poorly for patterns that are not in the training set and cannot achieve good generalisation performance [7]. The number of nodes in the hidden layer affects how well the network performs. The importance of network optimisation has been discussed in numerous ways but there is not mathematical expression till date [8-9]. For classification issues, it has been theoretically demonstrated that one hidden layer is sufficient to approximate any continuous function with arbitrary accuracy, providing there are enough hidden units [10].

The theoretical findings, however, do not provide guidance on how to select the appropriate number of hiddenlayer. Additionally, even though in theory one hidden layer might be sufficient, in practise it is recommended to use two hidden layers or more for quicker and more effective problem solving [11-13].

The use of more than two hidden layers is not justified theoretically [13].

Data normalization is applied to all the feature vectors in the data set first; creating a new training set and then training is commenced. Once the means and standard deviations are computed for each feature over a set of training data, they must be retained and used as weights in the final system design. Some research has reported the advantage of different type of data normalization e.g. Statistical or Z-Score Normalization, Min-Max Normalization, Median Normalization, Sigmoid Normalization, and Statistical Column Normalization [14].

A few scientists also provided information on the effectiveness of the activation function utilised in MLP. According to reports, "sigmoid" activation function outperformed "tansig," "sigmoid," and "purelin" activation functions [15]. Moreover, it has been observed that a two hidden layer neural network that uses the sigmoid function in every layer can communicate quickly with fewer epochs.

For methane and hydrogen they have low ignition energy so it can be ignite by sunlight. Therefore, it is crucial to detect flammable gases like CH_4 and H_2 in mining environments and other industries in order to monitor the environment, ensure worker safety, and reduce the risk of accidents. The majority of gas recognition systems are PC-based and therefore not portable. To eliminate every flaw in the cumbersome PC-based gas pattern recognizer that can save the lives of workers in mine camps and other industrial areas, an embedded system that automatically reads the sensor input from the array and computes the neural network output for the set of sensor values can be developed.

2. Methodology for the Classification of Gases using Neural Network

MLP is a supervised learning technique where the network has inputs vector (x_i), target vector (t_i) and error (e_i). In MLP a numbers of artificial neuron are present with free parameters. For a neuron (k) the weights (w_{ki}), one bias (b_k), one summer ($\sum$), one activation function (f) and output (y_k). Mathematical the representation of a simple neuron having two inputs is

$$y_k = f\left(\sum_{i=1}^{n} w_{ki} x_i + b_k\right) \tag{1}$$

$$e_i = t_i - y_i \tag{2}$$

2.1 Activation Function

There are generally four type of activation function used in neural network. They are linear activation function, hard limit activation function, log-sigmoid activation function and tan-sigmoid activation function. In linear activation function, the output is linearly proportional to the inputs. If “n” is the input and “a” is the output. It can be represents as in fig. 1.

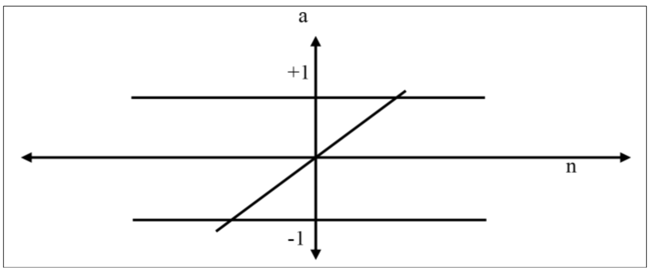

Fig 1: Linear Activation Function.

In log-sigmoid activation function, the output is lies between 0 to +1. The output is nonlinearly proportional to the input. This activation function is generally used in the hidden layer of a multi-layer perceptrons neural network. It can be represented as in fig. 2.

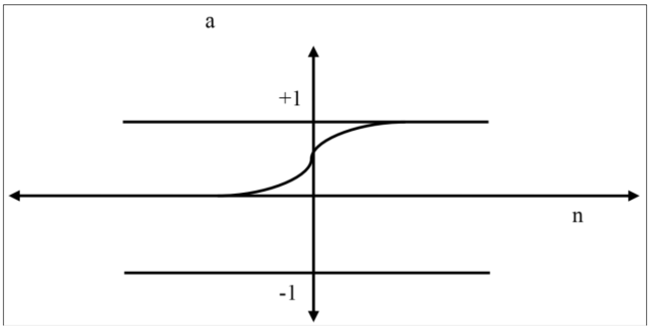

Fig 2: Log-Sigmoid Activation Function.

In tan-sigmoid activation function, the output is lies between -1 to +1. The output is exponentially proportional to the input. It can be represented as in fig. 3

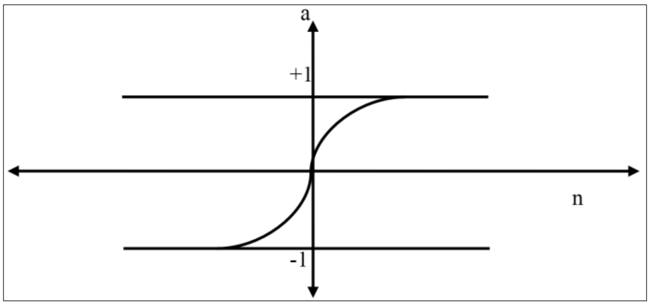

Fig 3: Tan-Sigmoid Activation Function

In hard–limit activation function, the output is either 0 or 1. Output is decided by using one threshold value. It can be represented as in fig. 4. The Hard-limit function is used in the perceptron network or in the output layer of the network. This function does not used in the hidden layer of the network.

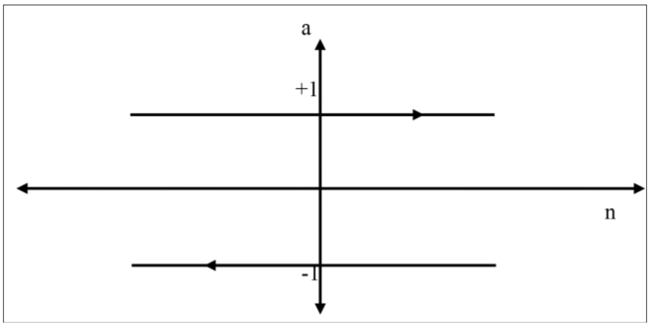

Fig 4: Hard-Limit Activation Function

2.2 Data normalization

The training of neural network can be made well-organized by performing certain pre-processing steps on the inputs of the network [3]. The normalization process for the real data inputs has provided suitable data for the training. Without normalization, training the neural networks would slower. Normalization can scale down the value of the data. Data normalization can speed up training time by starting the training process for each feature within the same scale. It is generally use for modelling application where the inputs are generally on widely different scales. There are three main data normalization. They are

1) **Statistical or Z-Score Normalization:** This technique uses the mean and standard deviation for each feature across a set of training data to normalize each input feature vector. The mean and standard deviation are computed for each feature. The mathematical expression is as follow:

$$y_n = \frac{x_n - \mu_n}{\sigma_n} \tag{3}$$

where y_n is the normalized value, x_n is raw data or real data, μ_n is the mean of the given set of data and σ_n is the standard deviation of the given set of data.

2) **Median Normalization:** The median method normalizes is simple and each sample is normalized by taking median of raw inputs. It is a useful normalization when there is a need to compute the ratio between two hybridized samples. Median doesn't influenced by the magnitude of extreme deviations of the data sample. It can be more useful when data set are widely distributed. The mathematical expression is as follow:

$$y_n = \frac{x_n}{median} \tag{4}$$

3) **Sigmoid Normalization:** The sigmoid normalization function is used to scale the samples in the range of 0 and 1 or -1 to +1. The output of the normalized data is depending on the transfer function used for normalization. It is useful when the estimated data are noisy. There are two kind of sigmoid normalization; they are tan-sigmoid and log-sigmoid. The mathematical expressions of these normalization are as follow:

Tan-sigmoid

$$y_n = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

Log-sigmoid

$$y_n = \frac{1}{e^x + e^{-x}} \tag{6}$$

2.3 Training Algorithm

There are eleven training algorithms namely, Broyden Fletcher Goldfard Shanno (BFGS), Conjugate Gradient with Powell/Beale Restarts (CGB), Fletcher-Powell Conjugate Gradient (CGF), Polak-Ribiére Conjugate Gradient (CGP), Gradient Descent with Momentum (GDM), Gradient Descent (GD), Gradient descent with momentum & adaptive (GDX), Levenberg-Marquardt (LM), One Step Secant (OSS),Resilient Back propagation (RP) and Scaled Conjugate Gradient (SCG).

2.4 Network with two hidden layer

The network with may have one single hidden layer or more hidden layer. With one hidden layer the mean square error is higher than the two hidden layer but higher the hidden layer may also take more memory. Therefore, the network with two hidden layer is consider for this study.

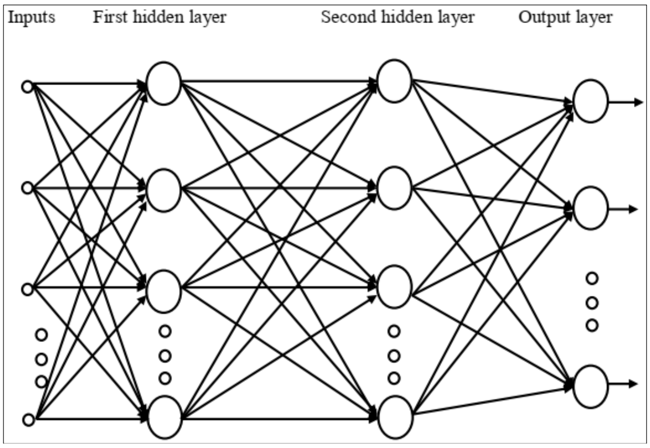

Fig 5: MLP with two hidden layer and one output layer.

3. Classification of CH_4 and H_2 using MATLAB

In MATLAB, classification of the CH_4 and H_2 gases can be done in two ways, one is by using nntool and another is by programming. First, the training data for CH_4 and H_2 classification are divided into two parts, one is for training set and another is for simulation to check the network output after training. A network with two hidden layer with five neuron in first hidden layer and eight neuron in second hidden and four output layer is consider for training with the raw date. The output layers are having four neurons because the data has two be categories with into four categories. They are CH_4 with 0.5%, CH_4 with 1%, H_2 with 0.5% and H_2 with 1%. For training, the input is matrix of 4X352 and target is a matrix of 4X352. For testing, the input is a matrix of 4X88 and target is a matrix of 4X88. The same network is train with different algorithm with log-sigmoid as the activations for all the neuron. The MSE outputs of the network are shown in fig. 6. In our analysis it is also seen that the mean square error for LM training algorithm is the minimum. Therefore, the LM training style is considered to be the best for our data set and is used for further analysis.

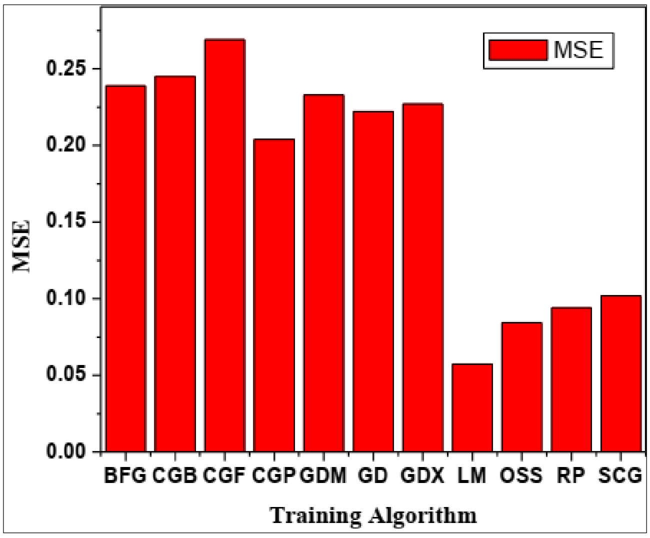

Fig 6: MSE for various learning algorithm.

3.1 Train with Raw Data

The network is train with raw data by varying the numbers of hidden neurons in first hidden layer and second hidden layers.

The mean square error of the networks with varied number of nodes in the hidden layers varies. The MSE of these networks with different numbers of hidden neuron are graphically represented as in fig. 7. From the graph, the optimum number of neurons in the two hidden layers in order to achieve best network performance is performed using LM training function is fourteen and twelve neurons in first and second hidden layers respectively with a minimum MSE of 0.00586.

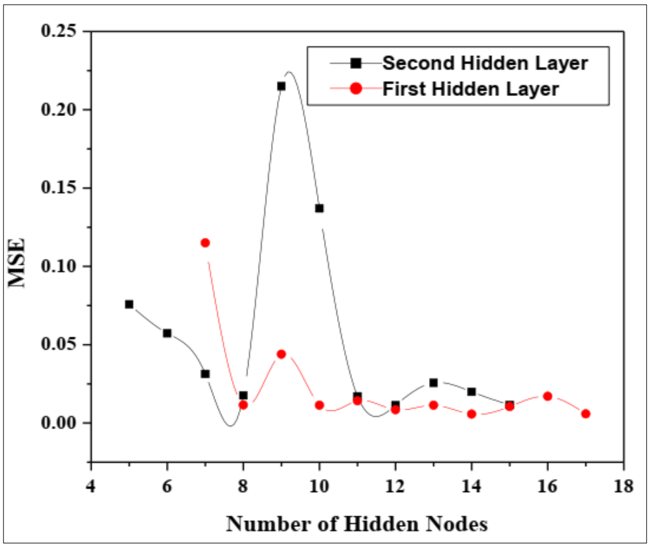

Fig 7: MSE Vs Hidden neurons with raw data training.

3.2 Train with Normalized Data

The input data are normalized the Z-Score normalization and used for training the networks. The training is done with different numbers of hidden neurons in first and second hidden layers.

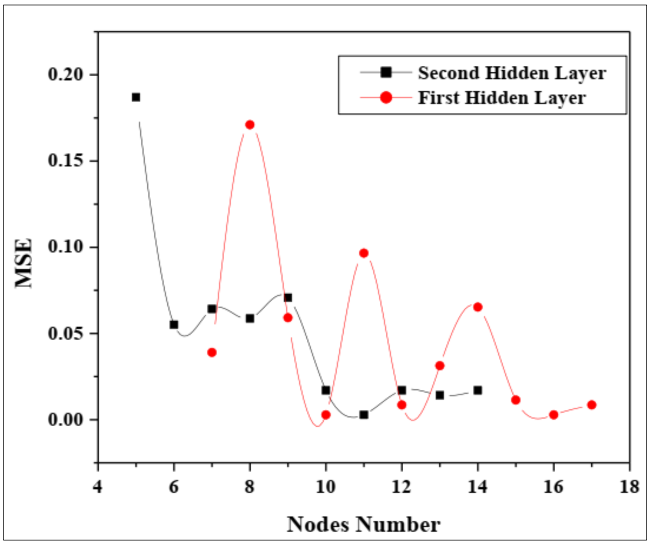

Fig 8: MSE Vs Hidden neurons with normalized data training.

Fig. 8 plots the MSE of the network with varied number of neurons in the first and second hidden layers. From the graph, the optimum number of neurons in the first and second hidden layers is ten and eleven respectively with a minimum MSE of 0.002.

4. Discussion and Conclusion

The training of the network is done with twelve different learning algorithms. The LM algorithm gives the best performance with minimum MSE amongst these algorithms. The network is train with different numbers of hidden neurons in the first and second hidden layer to find out the best network architecture. As it is observe that the MSE is lower as increase with numbers of hidden neuron then the MSE is increases with increase in hidden neurons after certain points. It is also observe that the training of the network with normalized data has less number of neurons in hidden layers with lesser MSE. This study clearly shows that classification of CH_4 and H_2 with different concentration level can be done using MLP. For raw data training, the numbers of neurons are fourteen and twelve neurons in first and second hidden layers respectively with a minimum MSE of 0.00586. For normalized data training the numbers of neurons are ten and eleven neurons in first and second hidden layers respectively with a minimum MSE of 0.002.

5. Reference

1. Ritaban Dutta KR, Kashwan M, Bhuyan EL, Hines JW. Gardner. Electronic nose based tea quality standardization. Neural Networks. 2003;16(5-6):847-853.
2. Dae-Sik Lee, Sang-Woo Ban, Minho Lee, Duk-Dong Lee. Micro gas sensor array with neural network for recognizing combustible leakage gases. IEEE Sensor Journal. 2005;5(3):530-536.
3. Anders H Andersen, William S Rayens, Yushu Liu, Charles D. Smith. Partial least squares for discrimination in fMRI data. Magnetic Resonance Imaging. 2012;30(3):446-452.
4. Sayago I, M del C Horrillo, Baluk S, Aleixandre M, Jesus Fernández M, Arés L, *et al*. Detection of toxic gases by a tin oxide multisensor. IEEE Sensors Journal. 2002;2(5):387-393.
5. Minghua Shi, Amine Bermak, Sofiane Brahim Belhouari, Philip CH. Chan. Gas identification based on committee machine for microelectronic gas sensor. IEEE Ransactions on Instrumentation and Measurement. 2006;55(5):1786-1793.
6. Shubhneesh Kumar, Maibam Sanju Meetei. Forecast and Analysis of Short Term Electric Load of New South Wales Region using ANN. International journal of engineering research & technology. 2014;3(6):1669-1671.
7. Kurt Hornik, Maxwell Stinchcombe, Halbert White. Multilayer feedforward networks are universal approximators. Neural Networks. 1989;2(5):359-366.
8. Mondal B, Meetei MS, Das J, Roy C Chaudhuri, Saha H. Quantitative recognition of flammable and toxic gases with artificial neural network using metal oxide gas sensors in embedded platform. Engineering Science and Technology, an International Journal. 2015;18(2):229-234.
9. Bebis G, Georgiopoulos M. Feed–forward neural network. IEEE Potentials. 1994;13(4):27-31.
10. Tsai J, Chou J, Liu T. Tuning the structure and parameters of a neural Network by using hybrid Taguchi-genetic algorithm. IEEE. 2006;17(1):69-80.
11. Stathakis D. How many hidden layers and nodes? International Journal of Remote Sensing. 2009;30(8):2133-2147.

12. Guang-Bin Huang. Learning capability and storage capacity of two-hidden-layer feed forward networks. IEEE Transactions on Neural Networks. 2003;14(2):274 - 281.
13. Gaurang Panchal, Amit Ganatra, Kosta YP, Devyani Panchal. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. Int. J Comput. Theory Eng. 2011;3(2):332-337.
14. Jayalakshmi T, Santhakumaran A. Statistical Normalization and Back Propagation for Classification. Int. J. Comput. Theory Eng. 2011;3(1):89-93.
15. Emad AM, Andrews Shenouda. A quantitative comparison of different mlp activation functions in classification. Springer-Verlag Berlin Heidelberg. 2006;3971:849-857.