Levenberg-Marquardt Deep Learning Algorithm for Sulfur Dioxide Prediction

Somia Asklany[1], Wahida Mansouri[1],Salwa Othmen[1]

[1]Northern Border University, Faculty of science and arts, Computers and information technology department, Turaif, K.S.A

Summary
Atmospheric pollutants play signification role in climate change as well as their great effect in human healthy. Prediction of such phenomena is very difficult due to the nonlinearity behavior of pollutant elements. Dynamic neural networks are good tools in dealing with such nonlinear problems that they can release the implicit dependencies in training data set through training algorithms for multilayer perceptron (MLP). The Levenberg-Marquardt algorithm (LM) is an epitome technique used to solve nonlinear problems. In this work MLP time series prediction with LM model will be build based on three years hourly data 2010:2012 cover great urban city Cairo, divided in three sets: 1500 target time steps for training (70%), over 450 target time steps for validation (20%) and over 225 target time steps for testing (10%). To avoid over fitting problem Levenberg-Marquardt algorithm stops training automatically when generalization stops improving, as indicated by an increase in the mean square error of the validation samples. Two performance measurements will be the methods of judgment the success of the proposed model: Mean Square Error (MSE) and relation coefficient (R). The proposed model is tested against recorded data set and proved superior.

Key words:
Levenberg marquard, neural networks, Deep Learning, Sulfur Dioxide prediction.

1. Introduction

Time series as a prediction technique collects past measurements of given variables variable and analysis it to discover the underlying relationships between them and finding out a descriptive model. This model will be applied to prognosticate future variable values. Applying time series in prediction has widespread in many areas of scientific research [1]. There are many types of time series that has some properties such as: nonlinearity, chaotic, non-stationary, and cyclic, from these types chaotic time series can be typically found in meteorological and climatological field especially in pollutants prediction [2]. To predict the behavior of chaotic time series which represent a nonlinear dynamic system, is a relatively complex problem that has drawn the attention and efforts of many scientists [3]. To overcome the limitation of time series Neural computing can be introduced to model nonlinear problem with more advantages compared with ordinary statistical methods. Artificial Neural Networks ANN can handle air dynamics phenomena comprise: multi seasons and large dataset [4]; [5]. MLP tree can give accurate result when forecasting Time-series [6]. ANN can develop a mathematical model for predicting daily concentrations of air pollution caused by the traffic in urban areas with high performance [7]. Two prediction systems in environmental field for two types of applications, one of them in the area of environmental protection used for air pollution prediction, and the other, in the area of hydrology, established for predicting of flood [8]. An intelligent system based on fuzzy rules used to predict rainfall events [9]. Three machine learning approaches used to capture temporal dependencies in the data of time series these approaches are: hybrid model based on long short-term memory (LSTM), multiple additive regression trees (MART) and deep feedforward neural network (DFNN) [10].
In this study Neural computing technique particularly (MLP) was introduced to create model based on (LM), this model used for SO2 prediction.

2. Area and Data

The data used in this study are three years data for Cairo urban city [2012–2014], the data collected from three locations cover mega city Cairo: Masr El Gadida, Fom El Kaleeg and El Gomhuria. The data include SO2 hourly observation data; SO2 is invisible and has a nasty, sharp smell. It reacts easily with other components to form harmful compounds such as sulfurous acid and sulfuric acid sulfate which harmfully affect human health as well as environment [11]. The data were divided in three sets; training, validation and testing data sets. Then the resultant model run over two days hidden data were chosen and selected randomly to check the behavior of the models. The data set was subjected to preparation process for handling noisy and missing data; this step is very important thus it leads to significantly improvement in the resulting model [12].

Manuscript received December 5, 2019
Manuscript revised December 20, 2019

3. Methodology

Neural Networks simulate the learning process in the human brain by amalgamating pattern recognition, reasoning and numerical computations [13]. The key feature of this paradigm is the innovative structure of the information processing system. It is consists of a considerable number of highly connected processing elements (neurons) working together to solve specific problems. The ANN can be applied to model large data sets with multi dimension s[14]. ANNs provide a method to solve non-linear problems that are difficult to solve by traditional methods. Although each neuron can be represented by a very simple structure, a collection of such elements establishes a tremendous processing power. Figure 1 is a schematic drawing of an individual neuron many of them amalgamating together to form the computing process.

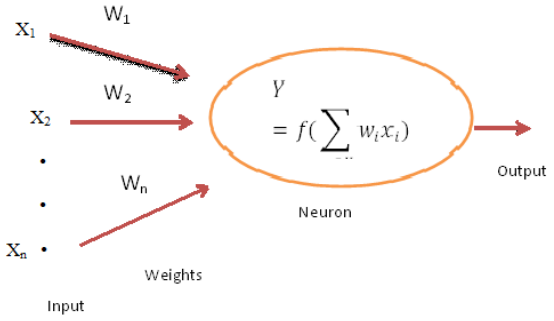

Fig. 1 The processing function of a single neuron

A neuron performs a computation with the inputs {x1. . . xn} to obtain an output value yi by activation function f as in the following equation :

$$y_i = f(\sum_{j=1}^{n} w_{ij}\, x_i) \qquad (1)$$

Where f(x) is the neuron function and the weights wij can be positive or negative, reproducing the so called excitatory or inhibitory character of neuron synapses, respectively.

To solve complex problems entails the need of complicated calculations, which can be performed by using networks with one or many extra layers between the input and the output knowing as layer multi-layer perceptron MLP [15]. Mathematically (MLP) is a pair (N,U), where N is a set of nodes and U is a set of processing units over N, which satisfies the following condition: Every node Xi ∈ N must be either an input or an output node of at least one processing unit in U.

Figure 2 structure an example of a multi-layer neural network with eight nodes
{x1 . . . x8} containing five processing units:
U1 = ({x1, x2, x3}, f1, {x4}),
U2 = ({x1, x2, x3}, f2, {x5}),
U3 = ({x1, x2, x3}, f3, {x6}),
U4 = ({x4, x5, x6}, f4, {x7}), and
U5 = ({x4, x5, x6}, f5, {x8}).

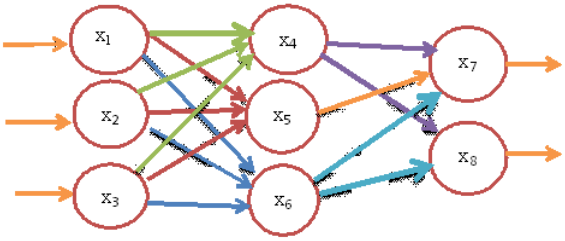

Fig. 2 feed forward Multi-layer Neural Network.

Deep learning (DL) is the novel kernel technique of MLP developed rapidly in nowadays. DL firstly was introduced by Hinton's deep auto-encoder (DAE) [16], which has multiple stacked restricted Boltzmann machines (RBMs). The learning process of DAE was divided into two different phases: first one , pre-training, which considered a kind of unsupervised learning using the gradient of network energy of RBMs, and the second one fine-tuning using the supervised learning: error-backpropagation (BP) [17].

One of the famous algorithms used to train a neural network is Levenberg-Marquardt (ML) algorithm; it has been designed to work specifically with loss functions which take the form of a sum of squared errors. It works without computing the exact Hessian matrix. Instead, it works with the gradient vector and the Jacobian matrix.

Consider a loss function which can be expressed as a sum of squared errors of the form

$$f = \sum_{i=1}^{m} e_i^{\,2} \qquad (2)$$

Here m is the number of instances in the data set.

Jacobian matrix can be defined by the loss function that containing the derivatives of the errors with respect to the parameters,

$$J_{ij} = \frac{\partial e_i}{\partial w_j} \qquad (3)$$

for i=1,…,m and j=1,…,n

Where m is the number of instances in the data set and n is the number of parameters in the neural network. Note that the size of the Jacobian matrix is m.n

The gradient vector of the loss function can be computed as:

$$\nabla f = 2J^T \cdot e \quad (4)$$

Where, e is the vector of all error terms.
Finally, we can approximate the Hessian matrix with the following expression.

$$Hf \approx 2J^T \cdot J + \lambda I \quad (5)$$

Here λ is a damping factor that ensures the positiveness of the Hessian and I is the identity matrix.
The next expression defines the parameters improvement process with the Levenberg-Marquardt algorithm

$$w^{(i+1)} = (w^i - (J^{(i)T} \cdot J^{(i)} + \lambda^{(i)} I)^{-1} \cdot \left(2J^{(i)T} \cdot e^{(i)}\right) \quad (6)$$

for i=0,1,…n
When the damping parameter λ is zero, this is just Newton's method, using the approximate Hessian matrix. On the other hand, when λ is large, this becomes gradient descent with a small training rate.
The parameter λ is initialized to be large so that first updates are small steps in the gradient descent direction. If any iteration happens to result in a failure, then λ is increased by some factor. Otherwise, as the loss decreases, λ is decreased, so that the Levenberg-Marquardt algorithm approaches the Newton method. This process typically accelerates the convergence to the minimum. In this paper LM algorithm had been used in training phase to different configurations MLP till reached minmum predicting error for the presented model.

4. Experiment

After data preprocessing step the optimal structure to the neural network should be established. To achieve this goal many configurations were created, trained and tested under try and error conception and by comparing the statistical parameters: mean square error (MSE) and correlation coefficient (R). MSE is a common criterion in optimization problems, it represent the average square difference between output and target.
The mathematical formula for computing MSE is:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(O - T)^2 \quad (7)$$

Where: n is number of data point, O is output and T is target value.
R, measures the strength and the direction of a linear relationship between any two variables: x and y as given in the equation 8.

$$R = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum y^2)}\sqrt{n(\sum y^2) - (\sum y)^2}} \quad (8)$$

Figure 3 represent R for the experiment.

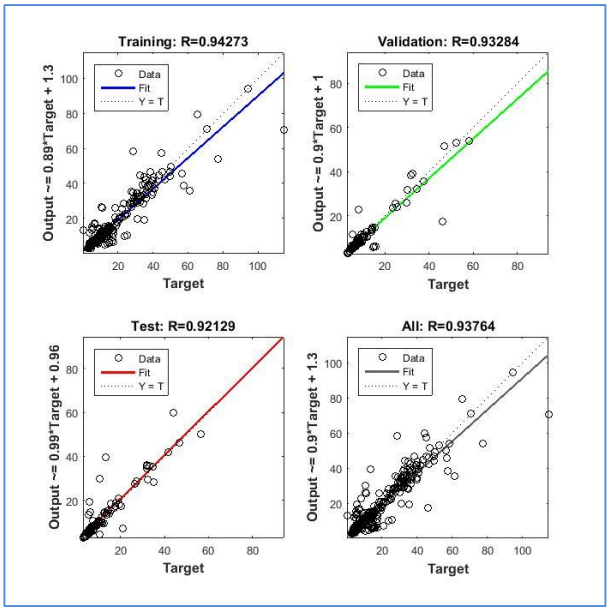

Fig.3 the correlation confident between output and Target

Sample results including the best performance shown in table 1.

Table 1: Sample statistical result

No. of hidden Neuron	No. of delay	MSE	R
50	3	9.5137e-0	0.9426
30	2	10. 3526e-0	0.9315
100	3	11. 3561e-0	0.9137

After the previous step, the following parameters can be determined: number of delay, number of neural network layers in the hidden layers, number of neurons in each layer and the type of the transfer function that in the hidden layers. The configured neural network surpassed other ones, is shown in figure 4.

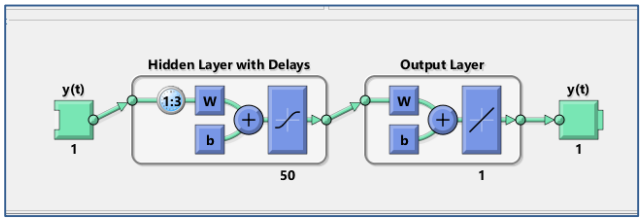

Fig. 2 The determined Neural Network.

In this configuration number of delay determined by 2 where number neurons was 50. The proposed configuration used to build model for predicting SO2 time series. The experiment was done under MATLAB environment [18].

The experiment can be represented by flowchart as in figure 5

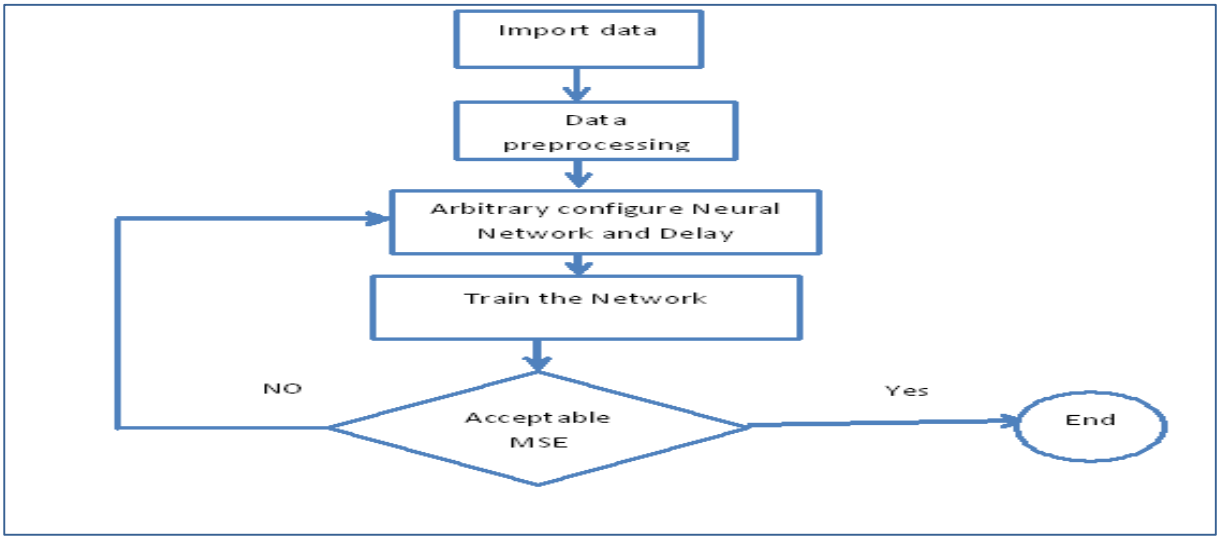

Fig.5 Experiment flowchart.

5. Results and Conclusion

In this paper Levenberg-Marquardt Deep Learning Algorithm was used to train MLP to get the best configuration for Neural Network by calculating the evaluation parameters: MSE and R. The resultant MLP used to build a model for predicting SO2 time series. The results prove high performance when testing against recorded data, figure 6 shows the time series response output target and error curves. The high accuracy of the model consequence the high correlation coefficient determined through the experiment. This high performance clearly appears when calculating MSE as shown in table 1. We can conclude that, the DL MLP with LM learning algorithm can build model able to capture temporal dependencies in SO2 time series data, which increased the accuracy of its forecasting. Therefore, this method can extend to be used in predicting the concentrations of others air pollutants.

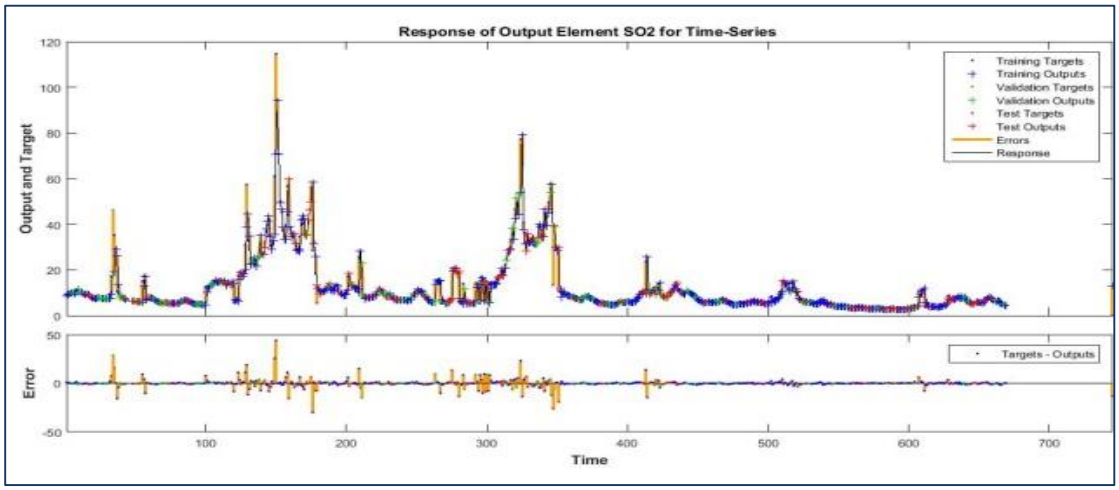

Fig. 6 Time series response

Acknowledgement

The author gratefully acknowledge the approval and the support of this research study by the grant no
SAT- 2018-3-9-F-7703 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A
.

References

[1] Nur Haizum Abd Rahman, Muhammad Hisyam Lee*, Mohd Talib Latif, Suhartono," Forecasting of Air Pollution Index with Artificial Neural Network", Jurnal Teknologi (Sciences & Engineering) 63:2 pp.59–64, 2013.

[2] Vlachas PR, Byeon W, Wan ZY, Sapsis TP, Koumoutsakos P. Data-Driven Forecasting of High Dimensional Chaotic Systems with Long-Short Term Memory Networks. 2018.

[3] S. A. Asklany, Khaled Elhelow, M .Abd El-WahabInternational ; On using Adaptive Hybrid Intelligent Systems in PM10 Prediction Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, vol. 6 Issue-4, pp. 54-59, 2016

[4] Zhiguo Zhang and Ye San .Adaptive wavelet neural network for hourly NOX and NO2 Concentrations. Winter Simulation Conference (WSC'04),vol. 2, 2004.

[5] Jose´ Luis Aznarte M., Jose´ Manuel Benı´tez Sa´nchez , Diego Nieto Lugilde , Concepcio´n de Linares Ferna´ndez, Consuelo Dı´az de la Guardia , Francisca Alba Sa´nchez Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models, Expert Systems with Applications., vol. 32,pp. 1218–1225,2007

[6] Yuehui Chen, Bo Yang and Ajith Abrahan.Time-series forecasting using flexible neural tree model. Information Sciences: an International Journal,vol 174, Issue 3-4. pp.219 – 235, 2005

[7] Bogdana VUJIĆ, Srđan VUKMIROVIĆ, Goran VUJIĆ, Nebojša JOVIČIĆ, Gordana JOVIČIĆ and Milun, BABIĆ. Experimental and Artificial Neural Network Approaches for forecasting of traffic air pollution in urban areas: the case of Subotica. Thermal Science, vol.14.pp.1-7,2010.

[8] Oprea, M. and Alexandra, M. Applying Artificial Neural Networks in Environmental Prediction Systems, Recent Advances in Electrical Engineering Proceedings of the 11th WSEAS international conference on Automation & information,pp.110-115. 2010.

[9] Somia A. Asklany a, Khaled Elhelow , Youssef I.K.and Abd El-wahab, M. Rainfall events prediction using rule-based fuzzy inference system. Atmospheric Research., vol. 101, pp. 228-236,2011.

[10] Hamed Karimian, Qi Li, Chunlin Wu, Yanlin Qi, Yuqin Mo, Gong Chen, Xianfeng Zhang, Sonali Sachdeva," Evaluation of Different Machine Learning Approaches to Forecasting PM2.5 Mass Concentrations", Aerosol and Air Quality Research, vol. 19, pp.1400–1410, 2019

[11] Kan, H.D., Wong, C.M., Vichit-Vadakan, N., Qian, Z.M. and Teams, P.P," Short-Term Association between Sulfur Dioxide and Daily Mortality",The Public Health and Air Pollution in Asia (PAPA) Study. Environ. Res., vol.110, pp. 258–264, 2010.

[12] A.,Jothimani,D., Big data: Challenges, opportunities and realities, In Singh, M.K., & Kumar, D.G. (Eds.), Effective Big Data Management and Opportunities for ImplementationPennsylvania, USA, IGI Global, pp. 1-24,, 2016.

[13] Stuart J. Russell and Peter Norvig. "Artificial Intelligence A Modern Approach", Prentice Hall Series in Artificial Intelligence, 2003.

[14] Gevrey, M., Dimopoulos, I. & Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological Modeling,, vol.160, pp. 249–264, 2003..

[15] M.E.Hamzehie ,S.Mazinani ,F.Davardoost ,A.Mokhtare, H.Najibi ,B.Van der Bruggen, S.Darvishmanesh," Developing a feed forward multilayer neural network model for prediction of CO2 solubility in blended aqueous amine solutions", Journal of Natural Gas Science and Engineering vol.21, pp. 19-25, 2014

[16] W. Tong, L. Li, X. Zhou, A. Hamilton, K. Zhang, " Deep learning PM 2.5 concentrations with bidirectional LSTM RNN ", Air Qual. Atmos. Health, vol. 12, no. 4, pp. 411-423, 2019.

[17] Hirata T, Kuremoto T, Obayashi M, Mabu S, Kobayashi K. Forecasting real time series data using deep belief net and reinforcement learning. Journal of Robotics, Network and Artificial Life.;4(4),pp.260-264, 2018

[18] MATLAB Version R2015a, The MathWorks Inc., Natick, Massachusetts, US. 2015

Somia Asklany was born in Egypt 1969, Assistant professor, Computer science department, Northern Border University King Saudi Arabia. Bs.c in Computer science and mathematics Minia University 1992, MSc in numerical modeling 2006 for application numerical prediction in studying the behavior of pollutant diffusion and Ph.D in Expert systems Cairo University 2011 in building expert system based on intelligent approaches: fuzzy logic and neural networks. Her research interest is artificial inelegance application in area of modelling and prediction techniques.

Wahida MANSOURI was born in Tunisia in 1983. She received her Ph. D degrees in Computer System Engineering, engineer's and master's degrees from the National Engineering School of Sfax (University of Sfax, Tunisia) in 2008 and 2010, respectively. Currently, she is member of the Electronics and Technologies of Information Lab of ENIS. Her research interests are communication networks, quality of services (QoS) and resource managements in next generation networks. Her current activities include scheduling and QoS provisioning for the integrated cellular networks and wireless local area networks (WLANs).

Salwa Othmen was born in Sfax in 1986. She received her Ph.D in Computer System Engineering from the National Engineering School of Sfax (ENIS) in January 2017. She is a member of Research Groups of SFAX Laboratory of Electronics and Information Technology (LETI) Laboratory since 2010. She obtained a Master Degree in Computer Science, Dedicated Computer System New Technologies (NTSID) from ENIS.. Her research activity is in the domain of optimizing and securing the

routing protocols in the multi-hop wireless networks (Ad Hoc, mesh network and cellular networks). She is currently an assistant professor at North Border University Saudi Arabia, where she teaching different courses related to computer science.