PAPER • **OPEN ACCESS**

Pollution level predictor using artificial neural networks trained with galactic swarm optimization algorithms

To cite this article: Nilay Nigam and D Geraldine Bessie Amali 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **263** 042093

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042093 doi:10.1088/1757-899X/263/4/042093

Pollution level predictor using artificial neural networks trained with galactic swarm optimization algorithm s

Nilay Nigam and D Geraldine Bessie Amali
School of Computer Science and Engineering, VIT University, Vellore-632014, India

E-mail: geraldine.amali@vit.ac.in

Abstract.Pollutant Level Predicator is a system which helps in predicting the amount of pollutants in a specific region. The system uses historic data in order to predict the value for the new input. The prediction system uses Artificial Neural Networks (ANN) trained with different optimization algorithms to classify the pollution level into several classes. This research paper assesses and analyses various techniques which can be used to predict the level of pollutant in Delhi. This study uses daily mean air temperature, relative humidity, wind speed and concentration of PM2.5 in Anand Vihar area of Delhi for a period of 2 years (2015 to 2016).Experimental results show that a ANN trained with Galactic swarm optimization algorithm produces a more accurate predication compared to other optimization algorithms.

1. Introduction

Air quality monitoring is an essential requirement in maintaining a pollution free environment especially in urban areas. Quality of air depends on a number of factors like time, location, vehicle concentration, time of the day and various other conditions. In India monitoring air quality is an important task as in today's world huge number of private vehicles as well as public vehicles are required for transportation resulting in high pollution. There are large number of factories around urban cities as well which produces harmful pollutants making air more hazardous to health. In this paper Delhi, capital of India, is taken under study as it faces a grave problem of high pollution which is rapidly growing. So this paper provides a comparison between several methods which exist to predict the level of pollution in a region.

Pollution level is determined by concentration of many pollutants like NO2, PM2.5, SO2, PM10 and other particles. But the most important one and having adverse effects to health is PM2.5[1]. Hence in this paper the level of PM2.5 of the particular area is taken into consideration along with the climatic factors - mean temperature, wind speed and relative humidity. The level of pollutant is classified into five classes- good, moderate, unhealthy, very unhealthy and severe depending on the amount of PM2.5 present.

This study uses daily mean air temperature, relative humidity, wind speed and concentration of PM2.5 in Anand Vihar area of Delhi fora period of 2 years (2015 to 2016). The data used was provided by Indian Meteorological Department (MMD) and U.S Embassy. To improve the accuracy of the model the data set is divided in the following categories- Training dataset which is used to train the model, Cross-validation dataset which is used to check validity of the result and Test set which is used to test the model.

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042093 doi:10.1088/1757-899X/263/4/042093

2. Literature Review

There have been several methods which are proposed and used over the years to predict the air pollution but they can be broadly classified into two categories- mechanistic and statistical models [1]. A mechanistic (deterministic) method shows how to numerically solve the set of differential equations. In this method requirement of large dataset is of minimal importance as the method requires in depth familiarity of causes of pollutants along with changes in the amount of emission from time to time and chemical composition of the pollutant. All the details regarding the emitter of pollutants as well as other related parameters is difficult to find [2]. The shortcomings and lack of available information increases the uncertainty of the result as the values which are used are either estimated or completely ignored. Mechanistic models are unable to give prediction for extreme events because of high complexity and uncertainty involved with turbulent flow but can predict frequent events with more accuracy. Insufficient knowledge about pollutant sources as well as emission inventories, can be the reason for errors in estimation of the air quality and high bias in mechanistic model. To overcome these shortcomings statistical model is used to predict air pollution level[3]. The use of statistical model to make predictions often result in higher accuracy as compared with mechanistic model because statistical model is suitable for description relationship between pollutants present in air and variables[4]. However, this model faces one drawback of not considering the physics underlying the data as a result the model which is developed is limited to only single area of study and is inapplicable to other areas. There was a comparative experiment conducted by Fernando et al comparing mechanistic and statistical model for daily forecasting of PM10 in 2005 at Central Phoenix station in Phoenix, Arizona, USA[6]. In this experiment a neural network and MODEL3-CMAQ were used as statistical and mechanistic models respectively.The performance of the statistical model was better and more accurate than mechanistic model.

Some of the most commonly used statistical approaches are:

- Classification and Regression Tress (CART): This technique uses a special software to identify variables which are strongly related with ambient pollution level. After identifying these variables they are used to predict pollution level on the basis of meteorological conditions and air quality[2].

- Regression analysis**:** In this method relation between the level of pollutants and climatic variables is determined by examining the historic data. It is a type of supervised learning method. The regression model can be employed for predicting the value or classifying the dataset. This method is simple to use does not work efficiently in the case of non-linear dataset.

- Artificial Neural Networks (ANN): This is another way for using the historical data apart from regression analysis in a manner which gives better accuracy for all kinds of data with less complexity. The Neural Nets made such that they can simulate the manner in which human brain recognizes patterns. ANN have been widely used for several applications and can be used for predicting air pollution as well [10].

- Support Vector Machines (SVM): It is a machine learning tool which is used for classification and can perform both linear and non-linear regression. It is advantageous to ANN in terms of its high learning capability using small training data but it is not used as much as ANN due its requirement of high number of variables as complex structure for choosing the similarity function.

Apart from choosing the model for prediction there has been increase in the usage of optimization techniques which are used with ANN to give better output. Genetic Algorithm, Simulated Annealing and Particle Swarm Optimization are some of the optimization technique.

- Genetic algorithm**:** It takes inspiration from evolution. The process of evolution starts from a population consisting of arbitrarily selected individuals and occurs in future generations [12]. Initially, fitness of every individual is evaluated in each generation then all the members of the current population are chosen on the basis of fitness function and then the new generation is generated on the basis of mutation of the previous generation. The modification process can be of simple recombination or mutation. In each subsequent iteration new population which is formed in previous step is used. Whenever either a maximum number of generations has been produced or a satisfactory fitness level is achieved for the population the entire process is said to be complete.The problem with it is that if the algorithm ends due to maximum limit in terms of generations then it is difficult to verify whether the solution is satisfactory or not.

- Simulated Annealing:This technique is based on the concept of annealing used in the field of metallurgy. In this process the concept of annealing is used which involves first heating the material then cooling it under controlled conditions which reduces its defects. Similar idea is implemented in algorithm as well by slowly decreasing the probability of accepting the worse solution as the solution space is traversed. This method is highly used as an alternative to Gradient Descent.

- Gradient Descent:It is a type of iterative optimization algorithm.By using this technique the local minimum for a function can be found by moving in a direction equivalent to the negative gradient. It is analogous to a situation where one tries to climb down a hill and in order to that the person takes the path which takes him to the least point in each iteration.

3. Use of ANN Model for Air Quality Prediction

The ANN used in this research paper consist of 3 layers.The first layer is for the input variables which are meteorological parameters. The hidden layer contains five neurons which is chosen on the basis of trial and error.The final layer which is the output layer comprises of the classes in which the data will be classified. In this paper the classes which are designed are-good, moderate, unhealthy, very unhealthy and severe to show quality of air. The transfer function which is used is the Hyperbolic tangent sigmoid function. The dataset is divided into three sections which are training dataset (60% of original dataset), cross- validation dataset (20% of original dataset) and testing dataset (20% of the original dataset).The performance of the network is evaluated by calculating the mean square error (MSE).

3.1 Basic Features of ANN

3.1.1 Multilayer Perceptron

Multilayer ANN or MLP has a layered architecture consisting of interconnected neurons. In general ANN contains three layers namely input layer, hidden layer and output layer, the neurons in each layer are connected with other neurons in the neighbouring layers. The result of a neuron depends on the connecting weights, these results acts as the input in the consecutive layer showing a direction for information processing. As a result MLP is also recognized as a feed-forward neural network. MLP is trained using the training dataset which helps the model to learn about the data on which it has to operate which means it helps the model to learn about the problem for which it is used.

3.1.2 Activation Function

The ANN consist of neurons which has a set signal function producing output which can go to number of other neurons in the network. In order to translate the input signal to output signal of every neuron a function called as activation function is used. Activation function can be classified in three classes: linear, threshold or sigmoidal. Linear function gives a resulting signal in the neuron which is in direct proportion with the activity level of the neuron. The Threshold function gives an output signal which stays constant till a certain level, then the output changes to a new level which stays constant until the activity level drops below the threshold. The Sigmoid function gives an output whose value changes continuously with the changes in the activity level although the variation is non-linear. It consist of two functions-Hyperbolic whose values varies from -1 to +1 and Logistic function whose value rages from 0 to +1. Sigmoidal transfer function is the most frequently used activation function.

3.1.3 Training Methods

The interconnections between the neurons is formed by the process of learning or training. Learning algorithm tells how the weights are to be adjusted so that the interconnections which are formed provides the best result. At first these weights are chosen at random but are then adjusted using the optimization techniques such that the result reaches at minimum value for the cost function.

4. Optimization Algorithms in training ANN

When we run a neural network it gives results which may or may not match with the expected output in such cases it is important to reduce the error and get a more accurate output. This process is called as optimization of neural network [12]. In this process a value of algorithm is found which gives the least error and minimizes the cost function. Optimization leads to a result which is more precise and makes best use of resources like memory and time as well as greatly increases the quality of the output.

There are a number of optimization techniques which are used such as Genetic Algorithm, Simulated Annealing and Gradient Descent. These large number of techniques often creates a dilemma of which algorithm is the best. There is no definite answer for it but researchers are working to find the best one. This paper compares the performance of three optimization techniques- Back Propagation, Particle Swarm Optimization and Galactic Swarm Optimization.

4.1 Back-Propagation

Back-propagation algorithm is a supervised learning algorithm. At first the input layer is fed with the corresponding inputs which are then multiplied with the weights of the respective interconnections between input and hidden layer and the product is passed on to the hidden layer. Inside the hidden layer the sum of the output from the input layer is calculated and a non-linear function processes that sum. Finally, output from the hidden layer is given to the output layer by multiplying it with the weights of the respective interconnections between hidden layer and output layer, the output layer processes this product and gives the final output of the model. The model for back-propagation neural network is shown in figure-1. The output from ANN is compared to the desired output and the error is calculated in the form of mean square error.This error is then fed back to ANN and the weights are adjusted accordingly to minimize the error in each subsequent iteration. The specifications of the ANN used in this paper is shown in Table 1.

Table 1. Specifications for the Back-propagation neural network (BP)

Sigmoid	Activation function used in the hidden layer
5	Neurons present at the hidden layer
Fmincg	Minimizing the cost function

After training ANN, the network is given test data in order to assess its performance. The networks which are trained properly give accurate result when presented with inputs which are not seen by the network before. The architecture of a feedforward neural network is shown in Figure 1.

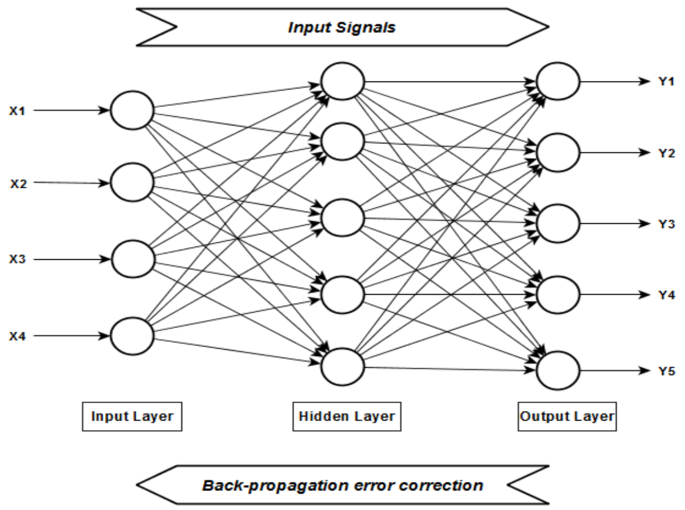

Figure 1. Back propagation model

4.2 Particle Swarm Optimization

PSO is a technique used for optimization whose basis is the movement and intelligence of swarms. There are a number of particles in swarm which move around the search space to look for the optimal solution[8].Every particle in the swarm behaves like a point present in the N-dimensional space which change their movements after carefully considering their own experiences as well as the experiences of the other particles present in the swarm. Every particle keeps track of the coordinates in the solution space which relates it to the optimal solution found so far by it.Pbest denotes the personal best value of each particle and Gbest is the global best which is defined as the best value obtained by the particles in the neighborhood of the particle under consideration.

The location of each particle is subject to change and depends on following factors - distance between the current position and gbest, current positions, current distance between the current position and pbest and current velocities. This information is given in equation 1

$$V_i^{k+1} = wV_i^k + c_1 rand_1(....)x\ (pbest_i - s_i^k) + c_2 rand_2(....)x(gbest - s_i^k) \quad (1)$$

Where,
rand: uniformly distributed random number between 0 and 1
v_i^k: velocity of particle i at iteration k
w: weighting function
s_i^k: current position of particle i at iteration k
gbest: gbest of the group and

$$w = wMax - [(wMax - wMain)xiter]/maxIter \quad (2)$$

Where,
wMin: final weight
wMax: initial weight
maxIter: maximum number of iteration

iter: current iteration number and
s_i^{k+1} is given as:
$s_i^{k+1} = s_i^k + v_i^{k+1}$
Specifications of ANN with PSO are given in table 2.

Table 2. Specification of PSO

Population size	20
c1 , c2	2.05
xmin, xmax	-1,1
The number of iteration	2000

4.3 Galactic Swarm Optimization

The GSO algorithm is highly inspired by the cosmos, its way of finding the optimal solution is similar to the movement of cosmological creations like stars and galaxies [6]. Inside large galaxies the stars are attracted towards the large masses, the same phenomenon is used in the algorithm. The working of the algorithm can be explained as follows. Initially, the solution which is better than the rest in solution space attracts all the individuals in the subpopulation using the PSO algorithm. Now the best solution of each sub-population represents their respective sub-population and the entire collection of sub-population is called a super swarm. Each individual in the super swarm comprises of the best solution found by each population and the movement of these individuals is based on the PSO algorithm[6]. The method explained above is generic and subjected to several variations which can be done to control the movement of swarm. In the GSO algorithm the swarm is denoted by set X consisting of D-tuples which contains the elements ($x^{(i)}_j \in RD$). Each set is filled with M partitions of size N and are denoted by X_i. Every element in X is randomly initialized from the search space given as $[xmin, xmax]^D$. Framework of the swarm can be given by:

$$X_i \subset X : i = 1, 2, \cdots, M$$
$$\mathbf{x}_j^{(i)} \in X_i : j = 1, 2, \cdots, N$$
$$X_i \cap X_j = \phi : \text{ if } i \neq j$$
$$\bigcup_{i=1}^{M} X_i = X$$

Where, X_i represents the swarm of size N.
$v^{(i)}_j$ gives the velocity and
$p^{(i)}_j$ gives the personal best associated with each particle $x^{(i)}_j$.

For each sub-swarm the PSO algorithm runs independently at the first level of clustering and runs for M times as the swarm X is subdivided into M groups. The most optimal solution found by the sub-swarm stochastically attracts the particles of each sub-swarm. All the particles are attracted towards the gbest related to sub-swarm as it lies in the vicinity of the local minimum.
The movements of a sub-swarmin X_i does not depend and does influence another sub-swarm $X_j$for $i \neq j$, resulting in an unaffected and comprehensive search. The search space is freely explored by each sub-swarm on its own and the search starts by calculating the position and velocity which are given in equation 3and 4

$$v^{(i)}_j \leftarrow \omega_1 v^{(i)} + c_1 r_1 (p^{(i)}_j - x^{(i)}_j) + c_2 r_2 (g^{(i)} - x^{(i)}_j) \quad (3)$$

$$x^{(i)}_j \leftarrow x^{(i)}_j + v^{(i)}_j \quad (4)$$

where, ω_1 is the initial weight and
r_1, r_2 are random numbers
which are given by

$$\omega_1 = 1 - k/L_1 + 1$$

$$r_i = U(-1,1)$$

in which k denotes an integer representing iteration value which varies from 0 to L_1.

In the next stage of clustering only the global best participates in order to form the superclusters. As a result, the collection of global bests from sub-swarms given by X_i forms a new super-swarm Y. This super-swarm make use of the most optimal solution which is previously computed by all of the sub-swarms thereby exploiting the previously computed information.

The first level of GSO algorithm can be primarily stated as an exploratory phase whereas the second phase is an exploitative phase. The specification of ANN with GSO is given in Table 3.

Table 3. Specification of GSO

Population size	5
Sub population	20
c1 , c2, c3, c4	2.05
xmin, xmax	-50, 50
vmin, vmax	-10, 10
Epoch number	10

5. Results and Discussion

The data set which is used was divided into training, test and cross-validation data. The training dataset comprises of 60% of the original dataset and 20% is selected as the test data. The evaluation is done on the basis ofMean Square Error and Root Mean Square Error which are one of the most frequently evaluation criteria and are given in equations 5 and 6.

$$\text{RMSE}=\sqrt{\sum_{i=1}^{n}\frac{(c_i - m_i)^2}{n}} \quad (5)$$

$$\text{MSE}=(\text{RMSE})^2 \quad (6)$$

Where,$C_i$is the optimal value which has been recorded
m_i is the value calculated by the model and
ngives the total data which is observed
The ideal value for RMSE is observed to be positive and close to zero.

Table 4. Comparison of the men square error obtained by backpropogation, PSO and GSO algorithms

Optimization Algorithm	RMSE	MSE
Back-Propagation	0.421452	0.177622
GSO	0.281760	0.079389
PSO	0.313366	0.098198

Table 4 shows the comparison of the mean square error obtained by backpropogation, PSO and GSO algorithms. The results which are obtained with this data set shows that the GSO gives the best performance with an accuracy of 92% and MSE as 0.07. PSO performs better than backpropogation with an accuracy of 90% and MSE 0.1. Microsoft Excel 2013 is used for handling data and Matlab 2010 is used for implementing ANN with BP and optimized using PSO, ANN with Back Propagation (BP), ANN with BP and optimized using GSO.

6. Conclusion

In this paper three models are compared in terms of their performance to predict the level of air pollution in Delhi using the information from Central Pollution Control Board and U.S Embassy Weather Station. GSO gives the best result in terms of accuracy and performance followed by PSO and backpropogation. By increasing the size of the original dataset and solving the issue of data instability the quality of prediction made by the network can be increased. The future work for this paper includes increasing the amount of input data to be fed in the network which can result in a better solution. In this paper the pollutant which is considered is PM2.5 other pollutants can also be taken into account as well.

References

[1] Hooyberghs J, Mensink C, Dumont G, Fierens F and Brasseur O 2005A neural network forecast for daily average PM10 concentrations in Belgium*Atmos. Environ***39** 3279–3289

[2] Hrust L, Klai´c Z B, Križan, J, Antoni´c O and Hercog P 2009 Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations *Atmos. Environ*. **43** 5588–5596

[3] Goyal P, Chan A T and Jaiswal N2006 Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmos. Environ. 2006***40** 2068–2077

[4] Jakeman A J, Simpson R W and Taylor J A1988Modeling distributions of air pollutant concentrations– III. The hybrid deterministic-statistical distribution approach*Atmos. Environ.***22** 163–174

[5] Chang J C and Hanna S R 2004 Air quality model performance evaluation. Meteorol.*Atmos. Phys.***87** 167–196

[6] Muthiah-Nakarajan V and Noel M M2016Galactic Swarm Optimization: A new global optimizationmetaheuristic inspired by galactic motion*Applied Soft Computing***38**771-787

[7] Vautard R, Builtjes P H J, Thunis P, Cuvelier C, Bedogni M, Bessagnet B, HonoreC, Moussiopoulos N, Pirovano G and Schaap M 2007 Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry transport models over four European cities within the CityDelta project.*Atmos. Environ*. **41** 173–188

[8] Amali G and Vijayan V 2015Accurate solution of benchmark linear programming problems using hybrid particle swarm optimization (PSO) algorithms*International journal of applied*

*engineering research***10**(4)9101-9110

[9] Borrego C, Monteiro A, Ferreira J, Miranda A I, Costa A M, Carvalho A C and Lopes M2008 Procedures for estimation of modelling uncertainty in air quality assessment*Environ. Int*. **34** 613–620

[10] Stern R, Builtjes P, Schaap M, Timmermans R, Vautard R, Hodzic A, Memmesheimer M, Feldmann H, Renner E and WolkeR 2008A model inter-comparison study focusing on episodes with elevated PM10 concentrations*Atmos. Environ*. **42** 4567–4588

[11] Fernando H J S, Mammarella M C, Grandoni C, Fedele P, di Marco R, Dimitrova R and Hyde P 2012Forecasting PM10 in metropolitan areas: Efficacy of neural networks*Environ. Pollut.*,**163** 62–67

[12] Amali G and Dinakaran M2016Solution of the nonlinear least squares problem using a new gradient based genetic algorithm*ARPN journal of engineering and applied sciences***11** 12876-12881