Genetic algorithm for the prediction of estimated scientific data: A case study of temperature measurements

G.Srinivas Reddy

Department of Physics and Chemistry

Mahatma Gandhi Institute of Technology, Hydearabad-500075, India

Abstract: The scientific data in any experiment is of statistical fluctuations which need to be analysed through the established techniques in order to extract the information. The proposed work in the paper establish to trace out outliers of the data in a statistical trend and the conformation of statistical gaussian distribution. Genetic algorithm is implemented to predict the data, and it is very interesting point to be noted that predicted data by GA technique is in well correlation with raw data with a insignificant statistical deviation which can be conformed through curve fitting technique. The outliers and deviation of Gaussian distribution is also studied by probability plottings of experimental and processed data.

Keywords: Genetic Algorithm, prediction, temperature, learning algorithm

1. Introduction

The most of the research has been of emerging interest in handling the data of large dimensions and analysis of multiple parameters. The processes of classification, prediction and statistical estimations have been generally criterion of the study in research problem [1,2]. The engineering problems in the fields of civil, electronics, information and image analysis and financial problems require the estimations in statistical manner [3]. The present research work in the paper belongs to the field of physics where te data processing is required and the analysis has to be made conveying the factors of influencing in the phenomenon under the investigation. The series of data points in the present experimentation conducted in the controlled conditions need to be analysed with help of algorithms attempted successfully in the earlier investigations. The work in the present paper is significantly involve the basic established stastical analysis for the conformation of reliability of the data generated in the experiment [4,5]. The most of the problems with higher dimensions and critical nature are of challenging issues in executing the research targets [5,6]. The critical analysis of any scientific domain is based on the non-formal techniques. The various parameters involving in

the phinominon leads to the critical situation of solution that has to be reached in the execution of problem [7]. The solution when it is being converged generally experience the difficulties in reaching the exact solutions. But the very general observations in the majority research analysis and simulations, every imlemented technique or algorithm has its owen defaults in the processes of updation at each iterations. The task of any research in achieving the output or results can be expected to be accurate and perfect. The degree of perfection and presence of minimal error can be depedent on the mathematical refinement based on the processes that have been proved to be efficient in the earlier studies [8]. The problems that have been under the investigation has to be carefully observed for the clarity to trace the relations among the various parameters involving in the system of execution [9,10]. Once the idea of connecting the parameters in a logical manner is successful and therefore there may be chance of reduction in the degree of difficulty and complexity in the process of execution that may lead to the solution of our interest. The recent advances in many research areas are deployment of the learning algorithms which intellectually estimate or predict targets that being modelled in framing the problems [11]. The very general learning algorithms that are recently implemented in many fields and able to produce the outputs of reliability, have been Artificial neural networks, Genetic algorithms and particle swarm optimization algorithms.

There have been majority of research areas in the scientific or engineering domains and also in the medical and financial sectors where statistics is required in processing for extraction information and the convey the interpretation of the results obtained in the research problem.

2. Genetic Algorithm

Genetic Algorithm is an algorithm developed based on the biological philosophy in mathematics. The coefficients of the techniques are indicated by chromosomes. The present work involve the experimental data of [x_1 x_2 … x_K], where k is length of the data. The same length of the chromosomes are modelled in the technique. The proposed technique is implemnted through data vector of experimental values. The process of executing Genetic Algorithm on the data in experiment will be initiated with population of matrix being initialised.

i. Initialization of population

The matrix of population in algorithm will be a set of the chromosomes of first generation in the method of searching process: The simulated chromosomes are given as $\{Y_i^j\}$ where i varies from 1 to K (lengthnof the data to be processed) and j varies from 1 to N where N is the dimension of population matrix.

ii. The fitness function and selection process

This function is of most importance and it is to be carefully modelled as per the requirements of the target going to be achieved. Let the data in the experiment be represented by vector x_i where i=1, …, K and let Y_i^j be initial population where j =1, …, N and i =1,…, K. Fittest value $f(Y^j)$ of individual chromosome of the population is simulated as given below

$$f(Y^j) = \frac{1}{\sum_{i=1} |Y_i^j - x_i|} \quad j=1, \ldots, N.$$

The chromosomes that are fit in the population function or considered for the next generations. In present work fitness values are directed to be greater than a threshold value ϕ and these chromosomes will act or function as a parents in each iterations. The threshold ϕ is set as Mean Squared Error.

iii. Cross-over

The process of interchange between the chromosomes will alter gene a local fitness value $f_i^j = |x_i - Y_i^j|$. The cross-over between two chromosomes will be a significant step of updation in the algorithm and local fitness is defined by $f_{av}^j = \frac{1}{K} \sum_{i=1}^{K} f_i^j$.

iv. Mutation

It is the process accomplishe for the substitution of random numbers to avoide local traps that is experienced by objective function. The mutation rate is fixed to 0.04.

3. Methods and measurements

The statical error generated in every scientific experimentation is of vital significance. The observations or the data obtained in the experiment are with a general deviations as the influence of out side environment and also with a manual erros that are generally committed in the physical process of handling instruments. The stastical deviations encountered in the process of experimentation must be processed and need to be reduced in order to acquire the precision in the information which is going to be concluded in the investigation. The prediction is the popular technique for analyzing the paprametrs and to the statistical outliers produced in the process of measurements. The present work implemented in the paper involve the data of temperature that have been acquired by a scientific instruments (3 thermometers of different types) for three days in the same location in the time period 5 to 7 hours in controlled environment have been considered. The average data given by the 3 types of thermometers at every 25 minutes represents a data point in the experimrnt. The total

number of data points obtained in the experiment are 150. The sampling interval of data is 20 minutes. The statistical distributions are simulated on the data samples to see the pattern that the majority data points obey. The various established techniques in statical tool are implemented totrace out the statistical distribution pattern and outliers of the data observed. The data is considered as time series data which is applied to genetic algorithm. The process of prediction have been simulated on the raw data obtained in the experiment to see the performance of proposed Genetic algorithm on the data set. The quantile plot is attempted to see the comparision between theoretical values of the distribution which is obeyed and the experimental data values. The predicted data generated by Genetic algorithm and the raw data in the experiment are compared to see whether the deviations are observed. Statistical histograms are plotted for the both raw data and predicted data points. The curve fitting techniques are implanted to see the variations in both the data points of predicted and raw data.

4. Results and discussion

The raw data is processed for the prediction by the proposed Genetic algorithm and the data shown in the figures 1 and figure 2. It is conformed by the visual scale that no more change in trend followed by the raw data obtained in the experiment. Both the predicted data and raw data are distributed in the same way. A small deviation in the distribution of histogram is observed which is statistically insignificant. The both data exhibited identical trends. The curve fitting implemented on the both data indicates the level of mathematical refinement made through the parameter of adjusted r square. The predicted data was fit to straight line by 0.04 times original data which can treated as statistical insignificance. The mean square error observed in the simulation is found to be 0.08. It can be seen from the figure 3 that data is observed to be driven in perfect distribution gaussian distribution. All the observations that have been made indicates statistical verification of experimental data. The weightage of the data in the process of measurements can be evaluated through statistical plottings. The data which have been observed to deviated cannot be taken into account of the analysis in view of its weightage since it may been captured in abnormal conditions prediction by Genetic algorithm is found to more succeful as there is no change of statistical significance.

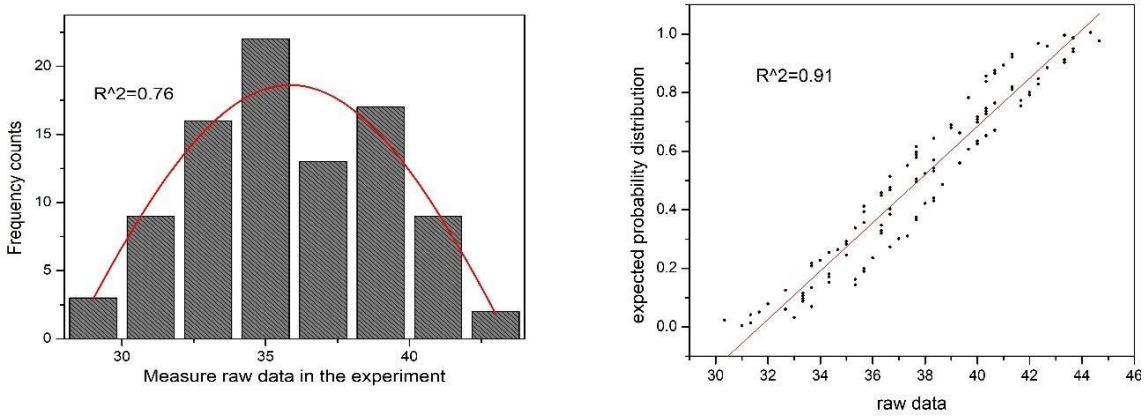

Figure 1(a) and (b)

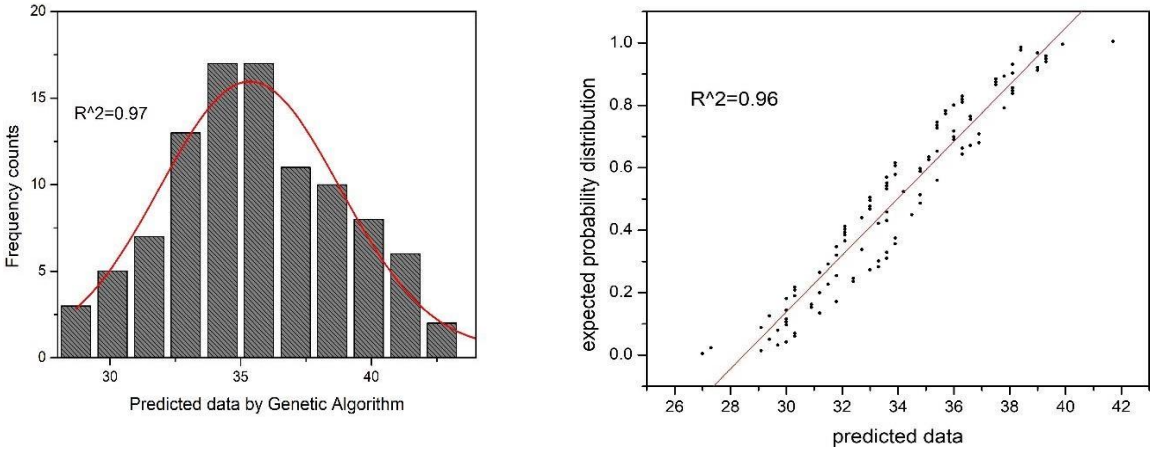

Figure 2(a) and (b)

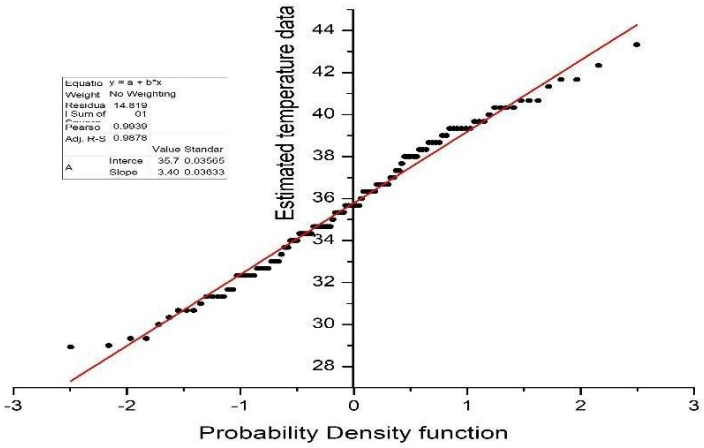

Figure 3. Probability of estimated data

5. Conclusions

The findings in the present work conveys the performance of Genetic algorithm on the data and the verification of statistical distribution followed by measured data in the experiment. One can get the statistical refinement of raw data through the implantation of genetic algorithm and established statistical techniques in order to achieve the processed data with a minimal error.

6. References

[1]. A.B.Brink, " Grey level thresholding of images using a correlation criterion, "*Pattern Recog. Lett.*, Vol. 9, pp.335-341, 1989.

.

[2]. Wilk, M. B. and Gnandaesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.

[3]. D.E. Goldberg, Genetic algorithm in search, optimization and machine learning, *Addison-Wesley, Reading,* MA, 1989.

[4]. Zhang Changli, Fang Junlong, Pan Wei. Research on Genetic Algorithm Trained Multi-layer Feed-orward Neural Network's Automatic Measuring to Tomato Maturity[J]. Transactions of the Chinese Society of Agricultural Engineering 2001;17(3);153-6

[5]. Ai Lirong, He Huacan. Summarise on Genetic Algorithms, Journal of Computer Application and Research, 1997, 14(4). 3-6

[6]. Xiao Yong, Chen Yiyun. Constructing Decision Trees by Using Genetic Algorithm, Journal of Computer Research and Development,1998,35(1).49-52

[7]. Mitchell M., An Introduction to Genetic Algorithm, A Bradford Book , MIT Press,1996,

[8]. M. Yu, N. Ena-Anant, A. Saudagar, L. Udpa, "Genetic algorithm approach to image segmentation using morphological operations", Iowa state university, Armes , IA, 1998

[9]. Schmidt M. and Stidsen T. [1997] , *Hybrid Systems : Genetic Algorithm , Neural Networks and Fuzzy Logic* , DAIMIIR

[10]. N.M. Razali and J. Geraghty, "Genetic algorithm performance with different selection strategiesin solving TSP," in *Proceedings of the World Congress on Engineering (WCE '11)*, vol. 2, pp. 1134–1139, London, UK, July 2011.

[11]. R.M. Haralick and L.G. Shapiro, "Image segmentation technique", *Computer vision, graphics and image processing, vol.* 29, 100-132, 1985