



Data Mining Formulation over Genetic Algorithm

Swati Jain, Vikas Kumar Jain¹, Sanjay Kumar,

Sunil Kumar Kashyap

Department of Computer Science, Kalinga University, Raipur, Chhattisgarh (India)

¹*Department of Basic Sciences, Government Engineering College, Raipur, Chhattisgarh (India)*

ABSTRACT

Data Mining is studied over Genetic Algorithm in this paper. The data lies with selection, crossover and mutation defined genetically. Data and information are the two different words with the same dimension. Data is the set of information. An information is an element but data is always referred a set. This paper deals with the data analysis by Genetic Algorithm.

Keywords: Data Mining, Genetic Algorithm.

I. INTRODUCTION

Modern world is influenced with the optimization; in this direction Genetic Algorithm has been proving its applicability. Genetic Algorithm is the technique of searching and selecting the best fit element as the best gene for the future generation efficiently. The Selection, Crossover and Mutation are the three foundation of any GA. The best gene to be selected, for the suitable operation and the proper mutation, it is all for the desired optimization of the objective function. The formation of the fitness function is a challenging task. If it is formed then its optimization is required. It cannot be obtained by only generalization but by the justified algorithm. Only with best gene, we can form a better population. GA generalizes this task up to the level of expectation. Goldberg [1] used GA first time to study the population as search, optimization and learning in the year 1989. In 1994, Srinivas et.al. [2] presented a survey on GA. Guofeng [3] proposed the acquirement of knowledge on data mining in 2003. It was the entry of data mining study via GA. In 2005, Zhou's [4] contribution is also noteworthy as per the composition the study over Neural Network and GA for forecasting the traffic flow. Yu [5] obtained a prediction by the modeling of NN and variants of other networks in 2005. Later, there were several applications came in the existence. In 2012, Choudhary et. al. [6] demonstrated GA over soft computing and presented a model for evaluating the performances of academic employees.

2. Generalized Genetic Algorithm: It is the process for generating the best variable for the next operation. The data of student's academic performance is studied as the application of GA.

Let the random variables are:

$$x_1, x_2, x_3, \dots, x_n.$$

The corresponding weights are:

$$w_1, w_2, w_3, \dots, w_n.$$

The summation of the weights is:

$$\sum_{i=1}^n w_i = w_1 + w_2 + w_3 + \dots + w_n,$$

or,

$$\Sigma = w_1 + w_2 + w_3 + \dots + w_n.$$

Hence, the The Activation Function is:

$$f(\Sigma).$$

Next is the fitness function, which is required for executing the process.

Let S be a total number of samples, G be the global error, t_i be the time at i position. Hence the fitness function will be;

$$f = \frac{1}{G} = \frac{1}{\sum_{i=1}^s (t_i)^2}.$$

The crossover function is defined in the next section.

Let X_i, X_{i+1}^t be the pair before crossover, X_i^{t+1}, X_{i+1}^{t+1} be the pair after crossover, C_i be the random number of uniform distribution in [0, 1], then,

$$X_i^{t+1} = c_i \cdot X_i^t + (1 - c_i) \cdot X_{i+1}^t$$

$$X_{i+1}^{t+1} = (1 - c_i) \cdot X_i^t + c_i \cdot X_{i+1}^t.$$

The algorithm is given in the next section.

1. The given data.
2. The Set of Random Numbers.
3. The coding by real numbers by;

$$L = i \times s + s \times j,$$

Where,

i be the Input random number, s be the sample random number and j be the out random number.

4. $\min(f)$.
5. New Population Generation by;

$$X_i^{t+1} = c_i \cdot X_i^t + (1 - c_i) \cdot X_{i+1}^t$$

$$X_{i+1}^{t+1} = (1 - c_i) \cdot X_i^t + c_i \cdot X_{i+1}^t.$$

REFERENCES

1. Goldberg D. E., Genetic Algorithms: A Survey, IEEE Computer, 27(6), 1989, 17-26.
2. Sreenivasrao V., Yohannes G., Improving academic performance of students of Defence University based on data warehousing and data mining, Global J. Comput. Sci. Technol., 12(2), 2012, 201-209.
3. Guofeng Q., Acquirement of knowledge on data mining, Computer Engineer, 29(21), 2003, 20-22.
4. Zhou A., A genetic algorithm based Neural Network approach for Short-Term Traffic Forecasting, Advances in Neural Networks, 3498, 2005, 965-969.



5. Yu W., Predictive model based on improved BP neural networks and its application, Computer Measurement & Control, 13(1), 2005, 39-42.
6. Choudhary O. K., Khot P. G., Deshmukh K. C., Soft computing model for academic performance of teachers using fuzzy logic, Br. J. Appl. Sci. Technol. 2(2), 2012, 213-226.