International Journal of Advanced Research and Development
ISSN: 2455-4030
Impact Factor: RJIF 5.24
www.advancedjournal.com
Volume 3; Issue 2; March 2018; Page No. 67-72

Genetic algorithm with different crossover for new analysis model of students performance

Somnath Maji
Department of CSE, Bankura Unnayani Institute of Engineering, Bankura, West Bengal, India

Abstract
Genetic algorithm (GA) plays an important role, as heuristic search techniques for handling complex spaces, in many domains such as artificial intelligence, data analysis, engineering, robotic, etc. The genetic processes on the natural evolution principles of populations have been fairly successful at solving complex problems and produce optimal (near optimal) solution from generation to generation. This is applied in students' quantitative data analysis to identify the most impact factor in their performance in their curriculum. The results will help the students as well as educational institutions to improve the quality of teaching after evaluating the marks achieved by the students' in academic career. Also used some different crossover in Genetic Algorithm for analysis and comparative study of results. So proposed genetic algorithm with different crossover with new student analysis model to find the most impacting factor in students performance.

Keywords: students' performance, quantitative factors, genetic algorithm, different crossover, influencing parameter, students evaluation results

Introduction
Genetic Algorithms (GAs) are adaptive heuristic search algorithms inspired by Darwin's theory of evolution (Darwin, 1859). In these algorithms a population of candidate solutions to an optimization problem evolves toward better optimal/ near optimal solutions based on the natural principles of inheritance, mutation, natural selection, and recombination. The evolution starts from a population of random or heuristically generated individuals (solutions) and continues through multiple generations. In every evolutionary step or generation, the individuals in the current population are evaluated according to some predefined quality criterion referred as aptitude function or fitness (Mitchell, 1997). To form a new population, individuals are stochastically selected from the current population based on their fitness and modified (mutated or recombined) to form a new population.

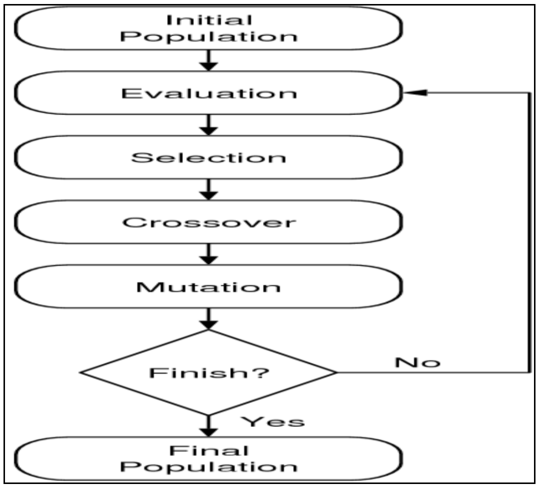

Fig 1

All of the genetic algorithms share a common structure that was developed by John Holland (Holland, 1975). In this structure, a population represents a group of possible solutions or chromosomes. In each iteration (generation), all of the population's chromosomes are evaluated using the fitness function. Then, the best chromosomes are selected from the actual population based on some selection criteria and a new population is generated based on the reproduction of selected individuals through genetic operators such as crossover (or recombination) and mutation.

Here we used different cross over method to compare the results. This generational process is repeated until a termination condition has been reached such as a fixed number of generations reached, allocated budget reached, an individual is found that satisfies minimum criteria, etc.

In today's educational system, performance in career growth is determined by the evaluation, assessment and examination achievements. The assessment is carried out by curricular activities such as class test, viva, seminar, assignments, general proficiency, attendance, lab work and finally semester exams. On considering the quantitative factors such as attendance, CGPA (cumulative), marks obtained in theoretical, mathematical, elective and departmental papers, knowledge in learning, understanding capability, communication etc. thus the computation is done by real genetic algorithm with different crossover is used to find the most influencing factor in those selected factors.

The remaining section of this paper is organized as follows, section 2 describes the several techniques used in evaluation of student's performance, section 3 discus about designing principles of various components in our student performance analysis model, section 4 describes about application of real genetic algorithm to find the most important quantitative feature, section 5 about working model of our Analysis tool and section 6 concludes the paper with remarkable outcome.

Literature Review

Here discuss about various techniques that are used to evaluate and predict the performance of employees, staffs, teachers and students and implementation of real genetic algorithm in bank applications. They are as follows. The most impact banking features in bankruptcy model are analyzed and results are produced. Deakin's bankruptcy model and its various features such as net income, cash assets, current assets, sales and total assets are considered. On application of Genetic algorithm they predicted the most important feature in this Deakin model is cash assets/ total assets. This ratio is crucial for predicting the bankruptcy with 96% accuracy (Martin *et al*, 2011) [1], similarly this approach is adapted here to find the most impacting parameter from the students results. A fuzzy Expert System tested with 20 student's marks obtained by semester-1 and semester-2 examinations, both inputs had same Triangular Membership Functions. The proposed a Fuzzy Expert System (FES) for student academic performance evaluation based on fuzzy logic techniques uses a suitable fuzzy inference mechanism and associated rule (Ramjeet *et al.*, 2011) [2]. The academic performances are taken and compare the results (performance) with existing statistical method. Similarly FES was developed for evaluation of teachers' performance in teaching activity is especially relevant for the academic institutions. It helps to define efficient plans to guarantee quality of teachers and the teaching learning process by considering students' feedback, result, students' attendance, teaching learning process and academic development of teachers (Chaudhari *et al*., 2012) [3].

The genetic operators on the tree representations are relatively straightforward. A mutation may modify the test attribute at a node or the class label at a leaf. A cross-over may substitute whole parts of a decision tree by parts of another decision tree. It is used to assist in the task of continuously monitoring a student's performance with reference to the possibility of passing the final exam. From these literature surveys, few researches have compared for analyzing the performance of student quantitative data and there is no considerable work on finding any equation model for to find the best parameter that influence the performance of the students. This research also analyzes the impact of quantitative parameters in student's academic performance. Using the different cross over we check the differences of performance results.

Different crossover in GA

Heuristic crossover (HX)

It is worth noting that the previous crossover operators did not exploit the distances between the cities (i.e. the length of the edges). In fact, it is a characteristic of the genetic approach to avoid any heuristic information about a specific application domain, apart from the overall evaluation or fitness of each chromosome. This characteristic explains the robustness of the genetic search and its wide applicability.

However, some researchers departed from this line of thinking and introduced domain-dependent heuristics into the genetic search, to create "hybrid" genetic algorithms. They have sacrificed robustness over a wide class of problems, for better performance on a specific problem. The heuristic crossover HX is an example of this approach and can be described as follows:

1. Pick a random starting city at one of the two parents.
2. Compare the edges leaving the current city in both parents and select the shorter edge.
3. If the shorter parental edge introduces a cycle in the partial tour, then extend the tour with a random edge that does not introduce a cycle.
4. Repeat Steps 2 and 3 until all cities are included in the tour.

Elitist Crossover (EX)

In the standard genetic algorithm, the selection process is always preceded by the crossover process. In the EX method, both processes are integrated. During the first step of the entire population is randomly shuffled. Then from each successive pair of parental vectors, two new vectors are created by crossover. From a 'family' created, two best vectors are singled out and implemented as offspring to the next population.

Application of elitist selection in the traditional way that is on the level of the entire population may often be the reason for the premature convergence of the algorithm. An EX elitist selection applied on the "family" level eliminates this danger according to the authors.

Masked Crossover (MX)

The MX operator uses a mask vector to determine which bits of which parent are inherited by the offspring. The first step is the duplication of the bits of the parents. The bits of the first parent are copied to the first offspring and, accordingly, of the second parent to the second offspring. In the second step, the offspring exchange bits among each other at those positions where the mask vectors of the parent were equal to 1, indicated domination of that parent at that position and the mask vectors of the other parent were equal to 0.

The mask vectors are initiated in P(0) randomly. During every iteration of GA, the mask vectors are inherited by each offspring from its parent. Then the mask vectors of the offspring as well as the parents undergo modification. The modification process is based on comparison of fitness of offspring and the parents. If good offspring are created, the masks of the parents do not need to be modified and the masks of the offspring may be very similar to those of the parents. In a situation where bad offspring were created the masks of the parents as well as of the offspring need to be modified.

Order based crossover (OBX)

The order based crossover operator selects at random several positions in one of the parent tours, and the order of the cities in the selected positions of this parent is imposed on the other parent to produce one child. The other child is generated in an analogous manner for the other parent

Modified Order Crossover (MOC)

A randomly chosen crossover point divides the parent strings in left and right substrings. The right substrings of the parent s1 and s2 are selected. After selection of cities the process is the same as the order crossover. Only difference is that instead of selecting random several positions in a parent tour all the

positions to the right of the randomly chosen crossover point are selected
For example with the following parents and crossover point
s1 = (1 2 3 4 □ 6 9 8 5 7) and
s2 = (2 1 9 8 □5 6 3 7 4)
After position selection
s1 = (1 2 * * * 9 8 * *) and
s2 = (2 1 * * * * 3 * 4)
Now obtain the generated pair of children as
b1 = (1 2 5 6 3 9 8 7 4) and
b2 = (2 1 6 9 8 5 3 7 4)
Clearly this method allows only the generation of valid strings.

Greedy subtour crossover

New crossover operator named `Greedy Subtour Crossover (GSX)' that acquires the longest possible sequence of parents' subtours. Using GSX the solution can pop up from local minima more effectively than by using simulated annealing (SA) methods. In the GSX, we use the path representation for a genetic coding. For example, chromosome g = (D, H, B, A, C, F, G, E) means that the salesperson visits towns D, H, B, A,.., E, successively, and returns to town D.
Suppose that chromosomes of parents are ga = (D, H, B, A, C, F, G, E) and gb = (B, C, D, G, H, F, E, A). First, choose one town at random. In this example, town C is chosen. Then, x = 4 and y = 1 because a4 = C and b1 = C respectively. Now the child g is (C). Next, pick up towns from the parents alternately. Begin with a3 (town A) because x <- 4 <- 1 = 3, and next is b2 (town D) because y <- 1<- 1 = 2. The child becomes g = (A, C, D).
In the same way, add a2 (town B), b3 (town G), a1 (town H), and the child becomes g = (H, B, A, C, D, G). Now the next town is b4 = H and town H has already appeared in the child (remember the salesperson may not visit the same town twice), so we can't add any more towns from parent gb. Therefore we add towns from parent ga. The next town is a0 = D, but D is already used. Thus we can't add towns from parent ga, either. Then, we add the rest of the towns, i.e., E and F, to the child in the random order.
Finally the child is g = (H, B, A, C, D, G, F, E).

Genetic approach for students performance analysis

In present's day Genetic algorithms are now widely applied in science and engineering as adaptive algorithms for solving practical problems. The proposed new model student analysis model using genetic algorithm(NSAMGA) is shown in below. The general acceptance is that GA is particularly suited to multidimensional global search problems where the search space potentially contains multiple local minima. Unlike other search methods, correlation between the search variables is not generally a problem. Genetic algorithm(GA) works modeling the parameters of a problem as real strings. The following figure depicts the various components of NSAMGA model.
The proposed model consists of starting edge from the student analyzing data with some constraints and ending edge by finding the most important parameter that is found by undergoing several processes as follows:

- All the performance analyzing the information in educational systems that are quantifiable such as attendance, internal assessment marks, project marks, previous semester marks, seminar, general proficiency, papers that are considered important in that course etc.
- Collecting those parameters, the equation construct using all those parameter values with weight values are designed. Also some constrained include here for range in parametric value.
- Then designing the equation, the real-coded genetic algorithm is applied.
- After that operations such as cross over (different crossover) and mutation are carried out.

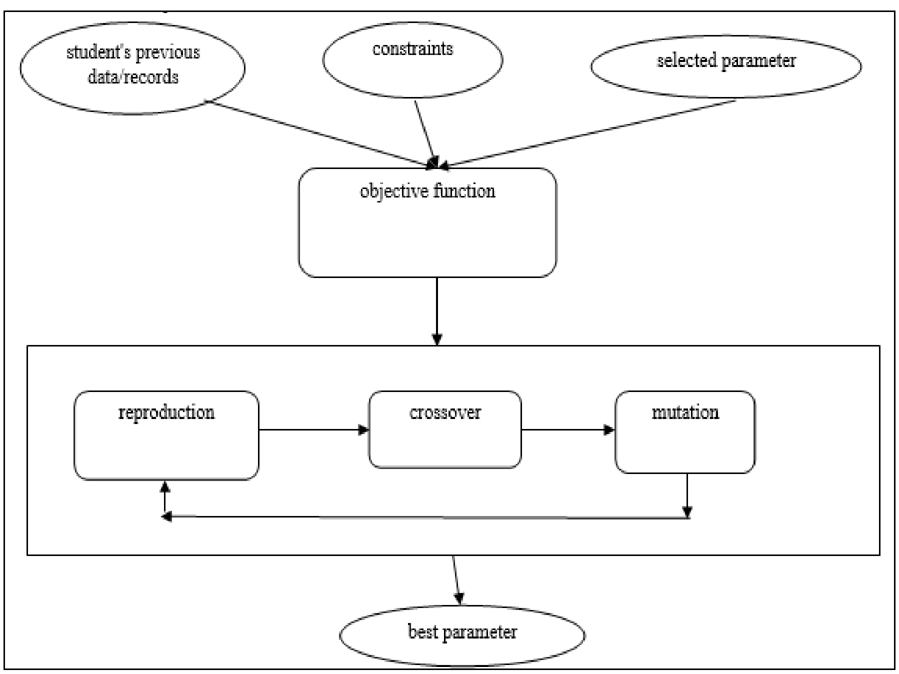

Fig 2: [Student analysis model using genetic algorithm (SAMGA)]

- The best parameter is identified, once analyzing all the parameters through real genetic algorithm. The quantitative factors that are considered into the analysis of students' performance are given below:

Table 1: Parameters

Variable	Description	Possible range of values
PSM	Previous semester marks	0-100
IA	Internal Assessment	0-25
L	Lab	0-50
P	Project	0-75
ASS	Assignment	0-5
ATT	Attendance	0-10
R	Record	0-5
GP	GP viva	0-5
T	Theoretical subjects	0-100
M	Mathematical subjects	0-100
E	Elective subjects	0-100
D	Departmental subjects	0-100
O	Other dept. subjects	0-100
PV	Overall Performance value	0-100

Most important parameter

The performance of analysis is concerned with finding most important attribute that affects the performance of student. As the fore mentioned properties for NSAMGA is designed using different Crossover and Mutation process. For this process we have selected the quantitative factors among the student in college/school. The real-valued genetic algorithm (RGA) with different crossover uses a real value as a parameter of the offspring in populations without performing coding and encoding process before calculates the fitness values of individuals. The performance analysis for NSAMGA model is given by the following equation:

Performance value (PV) = 0.08(IA) + 0.17(O) + 0.23(P) + 0.12(M) + 0. 18(T) + 0.22(PSM)
subject to constraints,

$$IA + O + P + M + T + PSM \leq 450$$

The working of real genetic algorithm with different crossover in performance analysis is as follows:

a. [Start] Generate random population of attributes as chromosomes
b. [Fitness] Evaluate the fitness f(x) of each chromosome x in the population
c. [New population] Create a new population of attributes by repeating following steps until the new population is complete
   1. [Selection] Select two parent chromosomes from a population according to their fitness (which satisfies the fitness function)(using Roulette wheel selection)
   2. [Crossover] with a crossover probability Pc (here used pc= 0.4) Crossover the parents to form a new offspring (children). For real values, Order Based Crossover (OBX) is performed. Also used heuristic cross over, greedy sub tour crossover, order cross over, masked crossover for comparison study.
   3. [Mutation] with a mutation probability Pm (here used pm= 0. 3) mutate new offspring at each locus (position in chromosome)
   4. [Accepting] Place new offspring in a new population

d. [Replace] Use new generated population for a further run of algorithm.
e. [Test] if the termination conditions are satisfied, stop, and returns the best parameter in current population
f. [loop] go to step b.

Initially a population of chromosomes, each of which represents a potential solution to the problem at hand, is generated randomly and each of them is evaluated by finding its fitness. The next generation of same size of population is created by selecting more fit individuals from this population and by applying genetic operators like different crossover and mutation to them. Mutation is an operator which creates a new individual by making a random change in the old one, whereas crossover creates new individuals by combining parts from multiple individuals. Classic mutation randomly alters a single gene, while crossover exchanges genetic material between two or more parents. Here used different crossover to make the solution of problem in different way. This completes one generation and after repeating this procedure until get the required solutions.

Results and Discussion

Table 2 illustrates the sample iterative process of the genetic algorithm to find the best parameter from the population. The selected iterated values are given in the Table 2. The fitness value is chosen from these values. According to that the influencing parameter can be identified. In this application, we have chosen the performance fitness value as > 2. Also used different cross over for comparative study of pv. Here we see that Genetic Algorithm with Order Based Crossover (OBX) is more effective for evaluating performance fitness compare to other crossover.

Table 2: Iterative process of the genetic algorithm with different crossover

Iteration No.	IA	O	P	M	T	PSM	Pv Using heuristic cross over	Pv using Order Based Crossover	Pv using greedy sub tour crossover	Pv using order cross over	Pv using masked crossover
3	2.00	1.00	1.00	1.00	1.00	1.00	1.17	1.20	1.06	1.11	1.07
6	1.00	1.00	1.00	1.00	1.00	1.00	1.46	1.50	1.25	1.36	1.11
8	1.70	1.00	1.00	1.00	1.00	1.00	1.54	1.70	1.36	1.65	1.59
10	1.90	1.00	1.00	1.00	1.00	1.00	1.66	1.90	1.77	1.47	1.29
12	1.00	2.00	1.00	1.00	1.00	1.00	1.05	1.15	1.02	1.01	1.09
14	1.00	4.00	1.00	1.00	1.00	7.00	1.89	2.45	2.02	2.09	2.03
16	1.00	4.00	1.00	1.00	1.00	1.00	1.65	1.75	1.66	1.69	1.87

18	1.00	3.00	1.00	1.00	1.00	1.00	1.87	2.05	1.98	1.78	1.69
20	1.00	10.00	2.00	1.00	1.00	1.00	1.99	2.35	2.09	1.39	1.47
22	1.00	1.00	4.00	1.00	1.00	1.00	1.34	1.75	1.88	1.58	1.58
26	1.00	1.00	8.00	1.00	1.00	1.00	2.12	2.75	2.23	2.14	2.08
28	1.00	1.00	10.00	1.00	1.00	1.00	2.36	3.25	2.28	2.26	2.18
30	1.00	1.00	1.00	2.00	1.00	8.00	1.98	2.40	1.96	1.99	1.66
32	1.00	1.00	1.00	1.00	1.00	9.00	2.15	2.70	2.01	2.09	2.07

The following diagram indicates the working of NSAMGA model for analyzing the student performance in mathematical, other department paper, and theoretical papers. The respective subject marks are loaded and their corresponding influence on students is shown in the below graph, the mathematical subject marks are taken whose maximum value is 100, given in y axis. The x axis represents the corresponding quantitative factors in terms of numbers. Thus each subject marks are analyzed from their semester values and results are produced separately to find the most influencing parameter as given below.
we have selected the 120 students of MCA(Master of Computer Application) departments and the results are used to analyze the most important parameter. This graph represents that parameter 1 and 4 have high impact on the performance of the students (1 indicates mathematical subjects and 4 is Theoretical subjects). Thus those two subjects have given high importance to reach the better performance. Next using fuzzy clustering technique to group the students on basis of their performance towards selected parameters.

Table 3

Reg. no	math	Other paper	theory
1	81	71	58
2	53	52	88
3	41	58	31
4	77	69	73
5	28	58	68
6	21	41	88
7	26	79	78
8	51	68	83

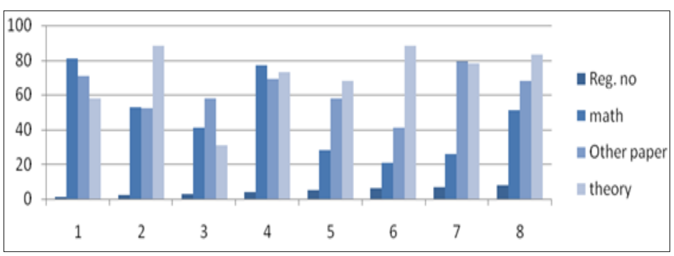

Fig 3

6. Conclusion

Thus my proposed genetic algorithm (GA) with different crossover for new model NSAMGA is successfully implemented to find the most influencing parameter that affects performance in the educational system. It will be a better solution for the classification, analyzing and evaluating the quantitative factors while course year and thereby it will predict the students' overall performance at the end of the semester examination and find results. This way of using Genetic Algorithm can be applied to any other areas to find the most influencing ratios or parameters to predict the performance effectively and efficiently and effectively. This research outlines the best features to develop a student performance model to analyze and can predict individual's performance efficiently. It will also act as a self-assessment and overall assessment tool for the students to know their position and area where to get improve. Thus we will provide the successful tool to predict the student's performance as well as group performance and overall performance using genetic algorithm. In future we can make this type of model more effective and efficient using some new crossover, new mutation and new selection procedure in genetic algorithm which can handle very large amount data to make the solution more and more optimal. And the effectiveness of this type of model towards student performance and overall performance of an institution will increase a lot.

References

1. Martin V, Prasanna Venkatesan, *et al*., To find the most impact financial features for bankruptcy model using genetic algorithm", International Conference on Advances in Engineering and Technology, ICAET-2011, 27-28.
2. Ramjeet Singh Yadav. Modeling Academic Performance Evaluation Using Soft Computing Techniques: A Fuzzy Logic Approach, International Journal on Computer Science and Engineering IJCSE, 2011, 3(2).
3. Chaudhari OK, Khot PG. Soft computing model for academic performance of teachers using fuzzy logic, British journal of applied science and technology. 2012; 2(2):213-226.
4. Oladokun VO. Predicting Students Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course, The Pacific Journal of Science and Technology. 2008; 9(1).
5. Osman Taylan, Bahattin Karagozog. An adaptive neuro-fuzzy model for prediction of student's academic performance computers & Industrial Engineering. 2009; 57:732-741
6. Emmanuel N Ogor. Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques published on Fourth Congress of Electronics, Robotics and Automotive Mechanics, 2009.
7. Tossapon boongoe, Qiang Shen, Chris Price. International Journal of Uncertainty, fuzzy qualitative link analysis for academic performance evaluation Fuzziness and Knowledge-Based Systems. 2011; 19(3):559-585.
8. Dimitris Kalles. Analyzing student performance in distance learning with genetic algorithms and decision trees, Hellenic Open University, Greece.
9. Harvinder Singh Atwal, Predicting Individual StudentPerformance by Mining Assessment and Exam

Data, 2010.
10. Aavo Luuk, Kersti Luuk, Predicting students' academic performance in Aviation College from their admission test results, 2011.
11. Mukta Paliwal, Usha, Kumar A. A study of academic performance of business school graduates using neural network and statistical techniques, E