I.J. Intelligent Systems and Applications, **2016, 9, 47-55**
Published Online September 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2016.09.06

Genetic Algorithm for Biomarker Search Problem and Class Prediction

Shabia Shabir Khan
Department of Computer Science, Research Scholar, University of Kashmir, Srinagar, India
E-mail: shabiakhan@gmail.com

S.M.K. Quadri and M.A. Peer
Department of Computer Science, Faculty of Computer Science, University of Kashmir, Srinagar, India

***Abstract*—**In the field of optimization, Genetic Algorithm that incorporates the process of evolution plays an important role in finding the best solution to a problem. One of the main tasks that arise in the medical field is to search a finite number of factors or features that actually affect or predict the survival of the patients especially with poor prognosis disease, thus helping them in early diagnosis. This paper discusses the various steps that are performed in genetic algorithm and how it is going to help in extracting knowledge out of high dimensional medical dataset. The more the attributes or features, the more difficult it is to correctly predict the class of that sample or instance. This is because of inefficient, useless, noisy attributes in the dataset. So, here the main aim is to search the features or genes that can strongly predict the class of subject (patient) i.e. healthy or cancerous and thus help in early detection and treatment.

***Index Terms*—**Genetic Algorithm (GA), Artificial Neural Network (ANN), Fitness Function, Feature Selection, Classification.

I. Introduction

Evolutionary computing inspired by natural selection and evolution process includes evolutionary strategies and genetic algorithms that can help in minimizing or maximizing the objective function to a large extent. The process goes through repeated evaluation of the objective function and each evaluation is followed by heuristic guideline check. Evolutionary strategies developed by Rechenberg helps in solving continuous optimization problems. Evolutionary programming uses FSM (finite state automata) to represent chromosomes that is capable of recognizing recurring patterns, even, odd or prime numbers. Apart from this, we have optimization technique of Genetic algorithm based on the concept of natural evolution that involves two main operations i.e. crossover and mutation used to find the best possible solution available. Further, genetic programming represents the solution by the parse trees of the computer programs that are evaluated on the basis of fitness value [1][2][3]. Taking into consideration an artificial intelligent technique of Genetic algorithm as feature selector, evaluation needs to be done by an intelligent evaluator probably a classifier, each time an optimal solution is obtained [12]. A classifier identifies the category or class to which an observation or instance belongs. This depends on the training, validation and testing set provided to the classifier wherein MSE (Mean Square Error) is calculated by comparing the observed value and the expected target value. The lower the error calculated, the better the optimized solution is. In this paper, we have used one of the best known classifiers or evaluators i.e. Multilayer Perceptron (Artificial Neural Network) which is mostly used in medical diagnostics [3][4][31].

II. Genetic Algorithm (GA) – Random-based, Derivative-free Evolutionary Algorithm

The working principle of Genetic Algorithm is based on the concept of natural evolution and science of Genetics provided by Charles Robert Darwin (1809 – 1882) and Gregor Mendel (1822- 1884). Genetic Algorithm is an optimization technique not actually looking for the best solution but looking for the good enough best solution rated against fitness criteria. So it avoids local optima and searches for global fitness. The algorithm takes huge search spaces navigating them so as to look for optimal combinations and provide the solutions we might not otherwise find in a life time. It is useful when search space is huge.

The Algorithm is considered as the variant of stochastic beam search bearing resemblance to process of natural selection where in successor state or the offspring of the state are generated by combining the two parent states rather than modifying the single state. The successor generated will populate the next generation depending upon its fitness value (fitness function). The four important factors that play an important role in the process and help in finding the hidden solution are: Fitness Function, Selection, Crossover and Mutation.

A well-known example of its implementation is the 8-Queen problem that can be easily solved wherein the goal is to place eight objects/queens in such a way that no queen attacks any other queen diagonally/ horizontally/vertically [4]. Genetics in biological world and Algorithm inspired from it has been discussed below:

A. Genetic field in Biological World:

As far as biological world is concerned we deal with two different make-up sections of cell/organism – Genotype (DNA sequence for genetic-makeup) and Phenotype (specific characteristic for physical make-up). In Genotype, genetic operators like mutation or crossover are being used for modification and recombination during reproduction. Whereas Phenotype selection operator is used to follow up the basic principle of Darwin i.e. "Survival of the fittest". In the field of genetics, there are three terms that are constantly being used by scientists and sometimes they are used interchangeably – Gene, Allele, Locus. For understanding that we need to go through some important biological terms:

i. Genome and Population:

Inside the nuclei of a cell in an organism, we have DNA molecule that is coiled around protein to form chromosome. The long double stranded DNA molecule contains four different bases i.e. Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) which form genetic code wherein the number and order determine whether you are a human or any other animal. The long molecules of DNA containing your genes are organized into pieces called chromosomes. Different species have different number of chromosomes. Humans having 46 chromosomes i.e. 23 pairs of chromosomes or two sets of 23 chromosomes, one set from each parent. The entire set of 23 chromosomes (i.e. all genes that define species) is called Genome. Set of competing Genomes or individuals is called Population.

ii. Gene and Allele:

Genes are small parts or functional units of this DNA molecule that code a specific feature e.g. eye color. Position of gene on chromosome is called locus. The gene(s) are being inherited by the child from its parent and is the reason behind looking like its parent. Value of a gene for specific feature like blue eye color or green eye color is called Allele and set of possible Alleles is called Gene.

iii. Genotype Operators:

The two important genetic operators are gene mutation and gene crossover or recombination.

a. Gene Mutation:

Gene Mutation is the permanent random change or alteration in the genetic DNA sequence of A, G, C, T. This may occur during DNA replication or external environmental changes. It is due to mutation that we have genetic diversity or variation within species.Fig.1 shows an example of mutation in a gene sequences.

Fig.1. Mutation in gene sequence

b. Gene Recombination in Gene Mapping (crossover):

Genes do not travel by themselves. They actually travel on chromosomes and as chromosomes undergo crossing over, genes from one chromosome actually swap position with gene on another chromosome at random points thus generating new variants by mixing existing genetic material. This is called Genetic Recombination. Fig2 below shows an example of crossover.

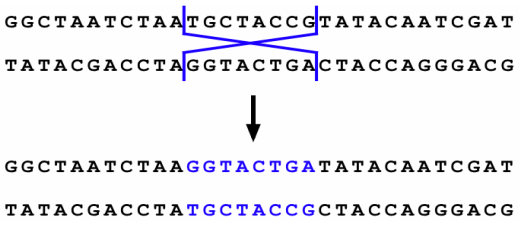


Fig.2. Crossover in chromosome

B. Genetic Algorithm inspired from biological genetics:

The artificial intelligent technique of Genetic Algorithm inspired by the Charles Darwin's principle of Natural Evolution helps in formulating the optimization strategies and thus solves the optimization issues in a wide variety of domains. Jhon Holland is considered to be the father of the Genetic Algorithm who proposed this noble idea in early 1970s. The evolutionary technique of genetic algorithms is an adaptive heuristic search algorithm based on principle of "natural selection" and "survival of the fittest" (Goldberg, 1989). It uses crossover and mutation operators in order to produce the offspring that form the next generation.

Under such an optimization technique, the population of problem solutions (usually variables represented as binary strings) is maintained in the form of individuals or chromosomes. These chromosomes consist of several genes which are actually the parameters of network. The problem solution strings are evaluated on the basis of an objective scoring function. At each step, following fitness evaluation, a new population of children chromosomes is produced for the next generation by applying rules or operations over the individuals or chromosomes from the current population/parents [1][2][4]

Gene->chromosomes->population->generation

Genetic Algorithm is an optimization technique that is based on the process of reproduction, genetic recombination and survival of the fittest. Here in our work we would consider the bit strings to be same as chromosomes in natural evolution. Further, the features shall represent an analogy to genes. The chromosomes are the solution candidates whose fitness or rank in the population is calculated using objective function or fitness function.

The three basic rules/ operations to create next generation from the current population are [4]:

i. Crossover: Combining chromosomes on the basis of crossover point (one-point or two-point or

uniform point) to produce new offspring.

ii. Mutation (boundary method): Selecting chromosome string and a random position over it to negate or alter the bit values.

iii. Selection (Roulette wheel method or Fitness Proportionate Selection): Selecting individuals or parents that are used to produce children for next generation. The selection process involves selecting individuals from one generation for their use in creating the next generation. We are using Roulette Wheel to determine the next chromosomes with randomly selected length wherein previous chromosomes are given chance to cooperate in the next generation so as to get stronger chromosomes. Roulette wheel method is based on the concept that the more the fitness, the more chance of survival of that individual exists [4].

Example: Example in Table 1 below represents a population of five individuals that are used for optimization and each individual or chromosome is a 10 bit string.

Table 1. Population of 5 individuals or 12-bit chromosomes

S.No.	Chromosome2	ValueBase-10	Var 'v'	FitFn(v)	Wheel Percentage
1	10101011100	2745	6.70	29.64	26.39941764
2	101101101001	2921	7.13	31.99	28.48916423
3	010000100011	1059	2.60	11.84	10.54913104
4	011010001100	1676	4.03	17.38	15.47381294
5	100000010000	2064	5.04	21.43	19.08847414
TOTAL				112.28	100

The steps involving crossover, mutation and selection, are repeated for several generations so as to find the best solution and the population of the strings with initial random parameters is created as candidates of best solution.

Following points detail out the example in Table 1:

1. The decimal representation of each binary chromosome is given in column labeled Value10.

   Chromosome: 11111111111112 = 409510

2. The column labelled Variable 'v' represents the normalized value calculated in the range [1: 10]. This is done by using the formula as:

   $$V = (\text{Value} / 4095) * 10$$

3. The fitness values are then calculated using fitness function of 'v'. Column labeled f (v) lists all the fitness values. The fitness function used here is :

   $$\text{Fitness function } f(V) = = (1/4)*V2 + (2 *V) +5$$

4. Sum total of all the fitness values can be used to calculate fitness percentage. This fitness percentage is actually the contribution of each chromosome in the roulette wheel. The more the fitness percentage, the stronger is the individual or chromosome and vice versa.

5. Here, chromosomes 101010111001 and 101101101001 are the fittest individuals because they have the largest share of roulette wheel. Others have smaller share and are weaker individuals. There is more probability for a selection point to be chosen from the largest share ones and thus probability of survival of fitter individual is more. These survived individuals participate in forming the next generation individuals.

6. Roulette wheel in Figure shows the fitness percent of each individual wherein each segment represents the area of particular individual. This Roulette wheel is spun 'n' number of times where 'n' is the size of population. There is greater chance of selecting the fitter chromosome for next generation, each time the wheel stops.

The various concerns to deal with in genetic algorithm include encoding of chromosomes, fitness or rank calculation, selection process, genetic operators and stopping criteria for genetic algorithm.

As per ranking of the chromosomes, the top n fittest chromosomes or individuals known by the name 'Elite' are being selected to move to the next or future generation i.e. Survival of a point in the parameter space is being decided by its fitness value. Those with higher fitness values are likely to survive and participate in the operations to produce better results.

III. Analogous Terms Used in Genetic Algorithm

In the field of genetics, there are several terms that are constantly being used by scientists and researchers. For understanding that we need to go through some important biological terms and relate them with the artificial intelligent technique of genetic algorithm [22] as follows:

A. Phenotype and Genotype in Genetic Algorithm:

Phenotype space in Genetic Algorithm is represented by the fitness values on the basis on which the chromosome is being selected for participation in the next generation. Genotype represents the population of

random individuals that are being evaluated for their participation.

As far as Solution Chromosome is concerned, genotype of an organism is often expressed using letters and the organism's phenotype is the visible expression of the genotype.

Genotype representation: (3021)10 = (0000101111001101)2

Figure 3 below shows the basic procedure:

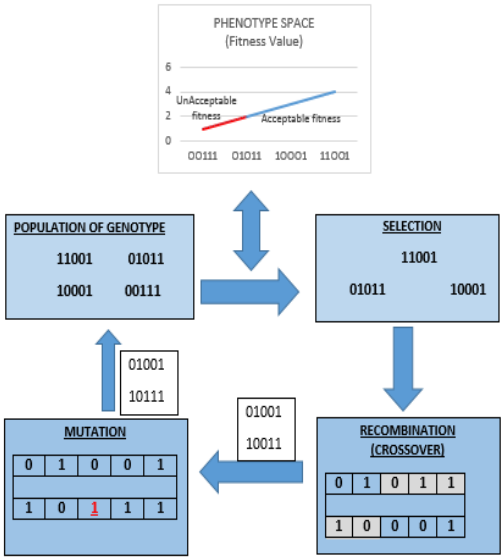


Fig.3. Genotype Space and Phenotype Space.

B. Crossover in Genetic Algorithm:

In Genetic Algorithm the concept of crossover plays a major role whereas Mutation has a minor role. E.g. We would like to apply crossover on parent stings having probability ranging between 0.5 to 1.0.The crossover site can be chosen randomly i.e. We can have one-point crossover or two point crossover.

C. Mutation in Genetic Algorithm

Mutation applied to allele or gene maintains the genetic diversity. Fig.3. above shows the mutated offspring after alteration or mutation in the third allele (bit complement).

IV. Algorithmic Steps

The basic idea behind Genetic Algorithm is to begin with a creation of Initial Population that consist of individuals or chromosome representing feasible solutions to the problem. Further the chromosomes (individuals) from the current population are selected based on evaluated fitness value and are directed to mutation and crossover processes so as to find the best solution chromosome for the next generation. The evaluation is followed for each iteration or generation trying to improve the fitness of Individuals and this continues to iterate until a particular criteria is met. The solution to the problem is the chromosome which is actually the collection of genes (features).

The steps are as follows:

Start

1. Create initial random population consisting of individuals represented by x.
2. Fitness evaluation of each individual using fitness function f (x)
3. Acceptable Solution found i.e. Termination/Stopping Criteria satisfied?
4. If Yes then Exit
5. Else
   i. Selection of parents or individuals according to fitness value including the best fitness elite children that automatically survive to the next generation some lower fitness value children or offspring.
   ii. Perform Recombination or Crossover to generate offspring
   iii. Perform Mutation of the offspring
   iv. Replace population by offspring to form the next generation.
   v. Goto 3

The parameters used in the algorithm are the mutation rate, crossover rate and the selection pressure (i.e. how much better the best individual is from the worst one).

Initially the population is considered to be in the range [0, 1]. As the generations increase, the chromosomes or individuals that form the population try to go nearer to the minimum point i.e. range [0,0]. The stopping or termination condition depends on the following points:

i. Number of generations specified,
ii. The amount of time specified for running the algorithm (time limit),
iii. The best point fitness value being less than or equal to some specified fitness value limit called fitness limit
iv. The relative change in the best point fitness value across generations is less than tolerance value specified.
v. There is no improvement in objective function over a specified period of time (stall time limit).

V. Optimization Algorithm for the Search of Strong Factors or Features

The implementation of Genetic Algorithm (optimization algorithm) begins with the random creation of population and followed by the fitness evaluation that calculates the rank of each chromosome. The chromosomes are probably in binary form wherein a '1' indicates presence of the gene i.e. the particular feature has been selected and a '0' indicates the absence of feature i.e. feature not to be considered in the chromosome evaluation. Following discusses the feature selection and implementation of genetic algorithm in

feature or gene selection:

A. Feature Selection

It is the process of obtaining a subset of optimal features for use in algorithm. Sometimes the high dimensional feature data set may affect the performance of the system due to some redundant or non-informative or irrelevant features or factors (often called as noise) [16]. To avoid such inefficiency and poor performance, we try to find out the best and smallest that represent the whole database even after removing some of the useless attributes or features. So, Feature Selector (Fs) acts as an operator to map m dimensional feature set to n dimensional feature set. The process is being used in order to get the filtered dataset with reduced dimension that improves the efficiency of the algorithm. The main aim of feature selection is to improve the performance of model by providing optimal subset of relevant features possible thus proving better prediction through any supervised classifier specified. As far as unsupervised technique is concerned, it can search better clusters using clustering technique. It helps in avoiding problem of overfitting and provide faster and cost efficient models [12][13]

Feature subset generation method has been classified under three main categories [24] i.e. Exhaustive Search method which is suitable for the small datasets as it is time consuming, however searches the best feature subset whereas Heuristic Search and Non-deterministic Search method uses evolutionary algorithm etc. (esp. genetic algorithm) to determine the feature subset and is usually applied over larger dataset.

B. Genetic Algorithm as a feature selector:

Various univariate techniques like filter, wrapper and embedded have some drawbacks which can be resolved by focusing on population based randomized technique of Genetic Algorithm along with classifiers to provide accurate solution [12][14][15][30].

Here, in our case study, we use genetic algorithm based feature selection technique using ANN based classification error as a fitness function thus optimal accuracy is obtained. Genetic Algorithm provides the different combinatorial set of features among which the best combination to achieve optimal accuracy is achieved. Genetic Algorithm has proved to be a very adaptive and efficient method of feature selection wherein the number of features is being reduced to an optimal number

$$Fs: Rr*m \rightarrow Rr*n$$

The major issue in genetic algorithm is the fitness function. This function is used to check how efficiently the selected subset of features (model generated solution) still represents the whole dataset. A classifier is applied every time a new combination of attributes is prepared to check the extent of how well that combination or feature subset is performing. This is done by comparing the error calculated by classifier for every feature subset i.e. the less the error, the better shall be the selected subset.

The algorithm ranks the function values in an order and passes the solution to next generation that has the best scores to produce those solution that'll prove to be better.

Feature selection is one of the important techniques that can be used in many application areas especially in the bioinformatics domain. The technique for feature selection using Genetic algorithm provides a filtered set for further analysis of useful information. Here, we shall focus on the supervised learning technique of Classification, wherein the class labels are already known.

VI. Related Works

Genetic Algorithm has been used in predicting the stock price index by selecting the efficient feature or attribute subset and optimizing the weight parameter between the layers in artificial neural network thus reducing complexity in high dimensional data and improving the learning technique of artificial neural network. This involved experimental comparison between the genetic algorithm and the conventional methods [16]. Similar research has been performed in Stock market prediction system wherein two methodologies have been performed i.e. data mining for extracting patterns and neural networks for extracting the valuable information from an large dataset, thus providing a reliable system [20]. Further, neural network has been used in the design of network topology by providing evolutionary-based algorithm that helps in simultaneously evolving the topology and optimizing the weights and number of neurons in neural network using evolutionary algorithm [21]. Apart from these fields, research work has been also done for facilitating better diagnosis in case of diseases like diabetes etc. To overcome the curse of dimensionality, Genetic algorithm has been used for feature reduction which, in turn, increases the accuracy in classification of patients with 'Major Depressive Disorder' treated with 'Repetitive Transcranial Magnetic Stimulation' [32].Genetic algorithm (GA) has been also used for feature selection and parameter optimization in breast cancer classification using resilient back-propagation neural network GAANN_RP for fine tuning of the weight and for determining the hidden node size [23].In case of diagnosis in diabetes, the work has been divided into two stages. In the first stage feature selection has been done using Genetic Algorithm and Neural Network classifier has been used for evaluation in the next stage [17][18]. Cardiac arrhythmia early detection is another field where the genetic algorithm has been used for optimization of learning rate and momentum in the neural network classifier. Here symmetric uncertainty provides reduced feature set and Simulated Annealing refines the population [19].

Traditionally, diagnosis depends on identifying some patterns from data that are obtained from human experiences. However this kind of diagnosis is prone to human error and is time consuming [23]. So, we need a better solution for overcoming the drawbacks and provide an efficient model. Genetic algorithm is used in attribute

selection for better prediction and reduction in the dimensionality of selected dataset.

The traditional method for diagnosing the disease relies on human experiences to identify the presence of certain pattern from the database. It is prone to human error, time consuming and labor intensive. Therefore, an evolutionary algorithm shall be used for filtering a feature subset and evaluated using a classifier to find out the best prediction system for diagnostics.

VII. Experiment over Pancreatic Tumor Dataset

Pancreatic cancer, a disease with poor prognosis and one of the major causes of death in the world, needs an early and efficient diagnostic system [25][26][28]. So our main aim is to find out the factors or features that actually affect or predict the survival of the tumor patient. This will help in early detection and treatment so as to increase the survival rate of the patients [23].

A. Recognizing patients as cancerous or heathy – A Case Study

i. Dataset Description:

Analysis of saliva supernatant from pancreatic cancer patients and healthy subjects- Oral fluid (saliva) meets the demand for non-invasive, accessible, and highly efficient diagnostic medium. Results provide insight into salivary biomarkers for detection of pancreatic cancer.

The experiment uses a salivary analytical dataset that consists of evaluated performance of gene features identified by gene symbol and reference transcript ID. The major aim was the early detection of pancreatic cancer. The experiment uses Affymetrix Human Genome U133 Plus 2.0 whole genome array so as to discover altered gene expression in saliva supernatant. The detection of salivary biomarkers can help in early diagnosis of pancreatic cancer specifically without the complication of chronic pancreatitis. The freely available dataset has been taken from Gene Expression Omnibus (GEO) is a database repository and can be referred for more information- GDS4100 [28].

The gene samples have been taken from pancreatic cancer patients and from healthy subjects. Total gene samples are 24 with 12 healthy tissues and 12 tumor tissues. The dataset (77×24) has been provided in two separate matrixes:

Matrix 'Inputs' – (76×24)
Matrix 'Targets'– (1×24)

The two class of patients are: Tumor and Healthy.

Table 2 below shows the class distribution of the dataset used:

Table 2. Class Distribution

Class	No. of Instances
Tumor	12
Healthy	12

ii. Fitness or Cost Function:

The implementation of Genetic algorithm for feature selection requires pre specification of maximum learning iterations and the population size for each iteration. Both are set to value 10. Further, the crossover and mutation percentages are specified, 0.7 and 0.3 respectively. The mutation rate is set to 0.1. The default evaluation method is Mean Squared Error (MSE), a statistical evaluation measure or estimator, is the mean or average of the squares of the errors or deviations wherein the error is the difference between the expected and observed value [3][4][23][29]. Eq. (1) below represents the formula for calculating MSE or simply Error 'E':

$$E = \frac{1}{n}\sum_{i}^{n}(T_i - O_i) \quad (1)$$

Further the cost or fitness value is calculated using the equation below:

$$\text{Fitness or Cost} = E^* (1 + P + SFR) \quad (2)$$

Where,

Parameter 'P' is the probability of the best chromosome or individuals selected compared to the average probability of selection of all chromosomes or individuals.

SFR is the ratio of selected features calculated as

i.e. SFR= No. of features selected/ Total No.of features)

VIII. Results

Selecting the classifier as ANN we first need to filter out the attribute set depending upon fitness value which is considered to be heart of Genetic Algorithm. This shall keep the balance between the applied classifier and the solution feature set. The experiment goes through an iterative process wherein each iteration results in a feature set that is being evaluated by the classifier. Each iteration searches the best cost value for the respective gene/feature set selected. Fig.4. below shows the screenshot of simulator calculating the best cost at 10th iteration with 74 attributes or gene set reduced to 40.

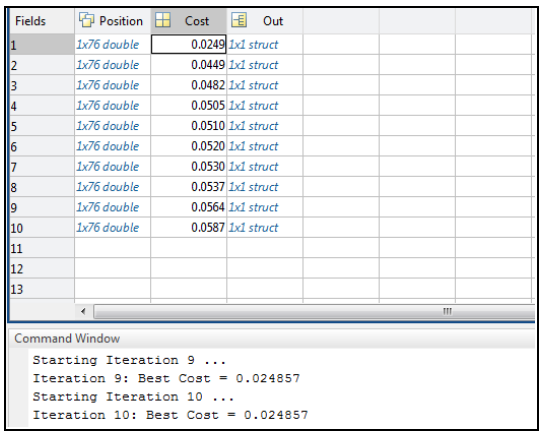


Fig.4. Cost function graph Pop structure (10*1) with 3 fields

The above screenshot (Fig.4) shows the population array with binary feature selection in 'position' variable, 'cost' calculated in cost variable and the 'out' variable containing the number of features selected. The best cost at 10th Iteration= 0.0249 which is known as the best Fitness Value.

The worst fitness Fitness Value recorded among all the iterations was: 0.7242.

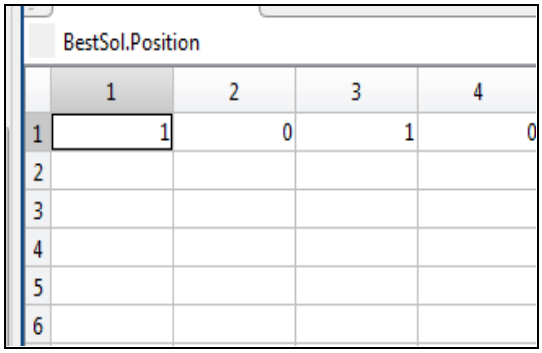


Fig.5. Structure with Value '1' indicating presence of feature and Value '0' indicating the absence of feature.

The best solution obtained with least cost value of 0.0249 is shown in Fig.5 above, wherein the binary value '1' is set for 1st and 3rd feature and binary value '0' is set for 2nd and 4th features, thus showing the selection of 1st and 3rd features among the strong cancer-predicting features.

Fig.6. below shows the graph for best cost value against each iteration. As the number of iterations increase, the cost value starts deteriorating, indicating the features set selected in the later iterations provide the lowest possible error and can strongly predict the status of the patient specially when it comes to poor prognostic diseases.

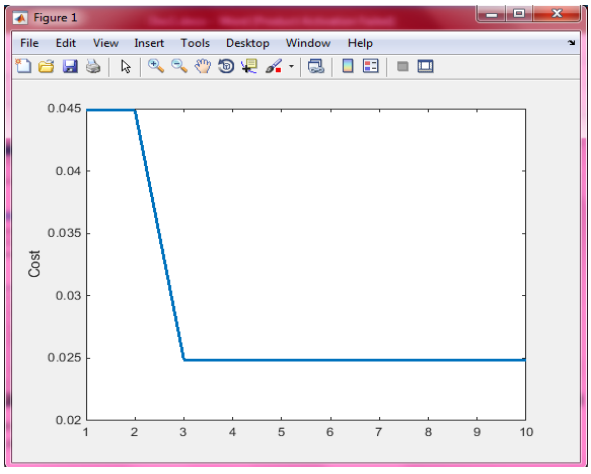


Fig.6. Best Cost Value (fitness value) against each iteration

Further, the Experimental results show that GA makes some of the gene features to outperform other genes or in other words we say the attributes or genes selected by Genetic Algorithm best predict or define the class of the patient i.e. either healthy or tumorous.

So the major focus for early diagnostics should be on feature set that has been filtered out by the genetic algorithm. Fig.7. below shows the graph for the best costs obtained in each iteration. The graph has same cost value in ending iterations.

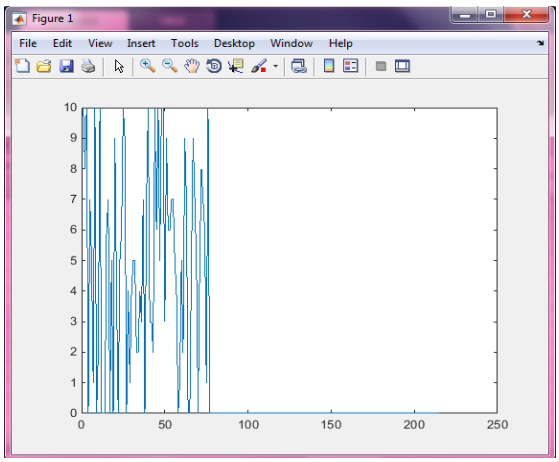


Fig.7. Frequency graph- Strength of the gene feature in prediction for each iteration.

Some of the genes or features that can strongly predict the class of the subject (patient) i.e. whether the subject is healthy or cancerous have been listed out below (Table3). These are the features that survived in each generation of genetic algorithm and participated in the next generation for finding out the best feature set (solution).

Table 3. List of gene features that strongly predict the class of subject (patient).

Gene Symbol	Gene Title
DDR1	discoidin domain receptor tyrosine kinase 1
HSPA6	heat shock 70kDa protein 6 (HSP70B')
PAX8	paired box 8
CCL5	chemokine (C-C motif) ligand 5
SCARB1	scavenger receptor class B, member 1
TTLL12	tubulin tyrosine ligase-like family, member 12
WFDC2	WAP four-disulfide core domain 2
MAPK1	mitogen-activated protein kinase 1
DPM1	dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit

IX. Conclusion

Genetic algorithm works on the principle of "survival of the fittest" and is based on wrapper approach of feature selection process. This concept can help in binary feature selection wherein the selected subset of features satisfy the fitness value specified and further participate in the next generation. The implementation of Genetic Algorithm along with classifier Artificial Neural Network filtered out some gene features that act as biomarkers and can strongly predict the class of the subject (patient) i.e. whether the subject is healthy or tumorous. The filtered gene or feature set could be worked upon for early diagnosis of pancreatic tumor.

REFERENCES

[1] Kampouropoulos, Konstantinos, et al. "A combined methodology of adaptive neuro-fuzzy inference system and genetic algorithm for short-term energy forecasting.Advances in Electrical and computer engineering". Volume 14, number 1 (2014).

[2] Tahmasebi. Peiman. and Ardeshir Hezarkhani. "A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation." *Computers & geosciences* 42 (2012): 18-27.

[3] Fakhreddine O. Karray, Clarence De Silva, "Soft Computing and Intelligent Systems Design- Theory, Tools and Applications". *Pearson Education*. 2009.

[4] S.N.Sivanandam, S.N.Deepa, "Principles of Soft Computing", *Wiley India Edition*,2007

[5] Hanafy, Tharwat OS. "A modified algorithm to model highly nonlinear system."*JAm Sci* 6.12 (2010): 747-759.

[6] Ge, Shuzhi Sam, and Cong Wang. "Adaptive neural control of uncertain MIMO nonlinear systems." *Neural Networks, IEEE Transactions on* 15.3 (2004): 674-692.

[7] Hanafy, Tharwat OS. "A modified algorithm to model highly nonlinear system."*JAm Sci* 6.12 (2010): 747-759.

[8] Goldberg D.E., "Genetic Algorithms in Search, Optimisation, and Machine Learning", Addison-Wesly, Reading, 1989.

[9] Michalewicz, Z., "Genetic Algorithms +Data Structures = Evolution Programs", Springer, 1996.

[10] Vose M.D., "The Simple Genetic Algorithm: Foundations and Theory (Complex Adaptive Systems)", Bradford Books, 1999.

[11] Matlab, "Global Optimization Toolbox User's Guide", The MathWorks, Inc, Revised 2015

[12] Yvan Saeys,Inaki Inza and Pedro Larranaga," A review of feature selection techniques in bioinformatics Bioinformatics" , BIOINFORMATICS REVIEW, Gene expression, Vol. 23 no. 19 2007, pages 2507–2517 , 2007

[13] Daelemans.W., et al., "Combined optimization of feature selection and algorithm parameter interaction in machine learning of language:A review of feature selection techniques",Proceedings of the 14th European Conference on Machine Learning (ECML-2003), pp. 84–95

[14] Li,T., et al. (2004) A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression.Bioinformatics, 20, 2429–2437

[15] Petricoin.E., et al. (2002) Use of proteomics patterns in serum to identify ovarian cancer. The Lancet, 359, 572–577.

[16] Kim, Kyoung-jae, and Ingoo Han. "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index." Expert systems with Applications 19.2 (2000): 125-132.

[17] Dilip Kumar Choubey, Sanchita Paul, Joy Bhattacharjee "Soft Computing Approaches for Diabetes Disease Diagnosis: A Survey". International Journal of Applied Engineering Research, Vol. 9, pp. 11715-11726, 2014

[18] Choubey, Dilip Kumar, and Sanchita Paul. "GA_MLP NN: A Hybrid Intelligent System for Diabetes Disease Diagnosis." (2016).

[19] V.S.R. Kumari, P.R. Kumar," Classification of cardiac arrhythmia using hybrid genetic algorithm optimisation for multi-layer perceptron neural network", International Journal of Biomedical Engineering and Technology, Volume 20, Issue 2, 2016

[20] Sudhakar, M., J. Albert Mayan, and N. Srinivasan. "Intelligent Data Prediction System Using Data Mining and Neural Networks." Proceedings of the International Conference on Soft Computing Systems. Springer India, 2016.

[21] Ahmadizar, Fardin, et al. "Artificial neural network development by means of a novel combination of grammatical evolution and genetic algorithm."Engineering Applications of Artificial Intelligence 39 (2015): 1-13.

[22] Melanie Mitchell(1996), An Introduction to Genetic Algorithms, A Bradford Book, The MIT Press, Cambridge, Massachusets Institute of Technology, 1996

[23] Ahmad, Fadzil, et al. "A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer." Pattern Analysis and Applications 18.4 (2015): 861-870.

[24] Nianyi Chen, Wencong Lu, Jie Yang, Guozheng Li, "Support Vector Machine in Chemistry", World Scientific ,Chap 4, pp.61, 2004

[25] Khan, Sheema, et al. "MicroRNA-145 targets MUC13 and suppresses growth and invasion of pancreatic cancer." Oncotarget 5.17 (2014): 7599.

[26] Moschopoulos, Charalampos, et al. "A genetic algorithm for pancreatic cancer diagnosis." Engineering Applications of Neural Networks. Springer Berlin Heidelberg, 2013. 222-230.

[27] Svetlana S. Aksenova, "Machine Learning with WEKA", WEKA Explorer Tutorial., 2004

[28] Zhang L, Farrell JJ, Zhou H, Elashoff D, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology,138(3):949-57, Mar 2010

[29] SM Kalami Heris, H Khaloozadeh , "Non-dominated sorting genetic filter a multi-objective evolutionary particle filter", Intelligent Systems (ICIS), Iranian Conference 2014

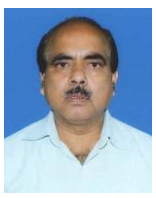
[30] Kumari, B., Swarnkar, T., "Filter versus Wrapper Feature Subset Selection in Large Dimensionality Micro array: A Review". IJCSIT, Vol.2 (3), pp. 1048-1053, 2011

[31] Amato, F.,Lopez, A., Maria, E.P.M.,Vanhara, P.,Hampl, A., Havel, J., "Artificial neural networks in medical diagnosis", J Appl Biomed, 11:47-58, 2013

[32] Erguzel, Turker Tekin, et al. "Feature Selection and Classification of Electroencephalographic Signals An Artificial Neural Network and Genetic Algorithm Based Approach." Clinical EEG and neuroscience 46.4 (2015): 321-326.

Authors' Profiles

Shabia Shabir Khan, Ph.D Research Scholar, Department of Computer Science, from University of Kashmir, Srinagar, Kashmir.



S.M.K.Quadri, Professor, Department of Computer Science, University of Kashmir, Srinagar.

M.A.Peer, Professor, Department of Computer Science, from University of Kashmir, Srinagar Kashmir.

How to cite this paper: Shabia Shabir Khan, S.M.K. Quadri, M.A. Peer, "Genetic Algorithm for Biomarker Search Problem and Class Prediction", International Journal of Intelligent Systems and Applications (IJISA), Vol.8, No.9, pp.47-55, 2016. DOI: 10.5815/ijisa.2016.09.06

