A BIRD'S VIEW ON RELATED LITERATURE ON GENETIC ALGORITHM

Srujana Rongali[1]

[1] Computer Science & Engineering Department,
GITAM University, Visakhapatnam, India.

Radhika Yalavarthi[2]

[2] Computer Science & Engineering Department,
GITAM University, Visakhapatnam, India.

Abstract

As quoted by Goldberg [1], Genetic algorithm (GA) is an adaptive optimization search algorithm aping the evolutionary ideas of natural selection. This Genetic Algorithm method is primarily applied haphazardly on an initial population and later all the individual chromosomes are appraised by a suitability function. This present paper is a review report on Genetic Algorithm and its functionaries on natural genetics and the evolutionary principle which was first proposed by Holland [2]. The GA is a simple but powerful tool for finding the global solution to an optimization problem.

Keywords: Chromosomes, Evolutionary Principle, Genetic Algorithm, Natural Genetics, Optimization

GENETIC ALGORITHM

Genetic algorithms (GA), is a generally considered as an adaptive optimization search as like Darwinian natural selection [3] and genetics in biological systems. This methodology is a promising alternative to conventional heuristic methods. The Genetic algorithms functions with a set of candidate solutions named as population. As like the Darwinian principle of 'survival of the fittest', the Genetic algorithm obtains the ideal solution after a cycle of iterative computations. Genetic algorithms transact with large search spaces competently, on a solution to the problem until satisfactory results are acquired by stimulating successive populations of substitute solutions that are represented by a chromosome. Allied with the characteristics of utilization and investigation search, this algorithm has less chance to get local optimal solution than other algorithms. A fitness function considers the quality of a solution in the evaluation step. In this method, crossover and alteration functions are considered as the main operators that haphazardly effect the fitness value. Chromosomes are selected for reproduction by assessing the fitness value. The fitter chromosomes have greater probabilities to be nominated into the recombination pool using the roulette wheel or the tournament selection approaches.
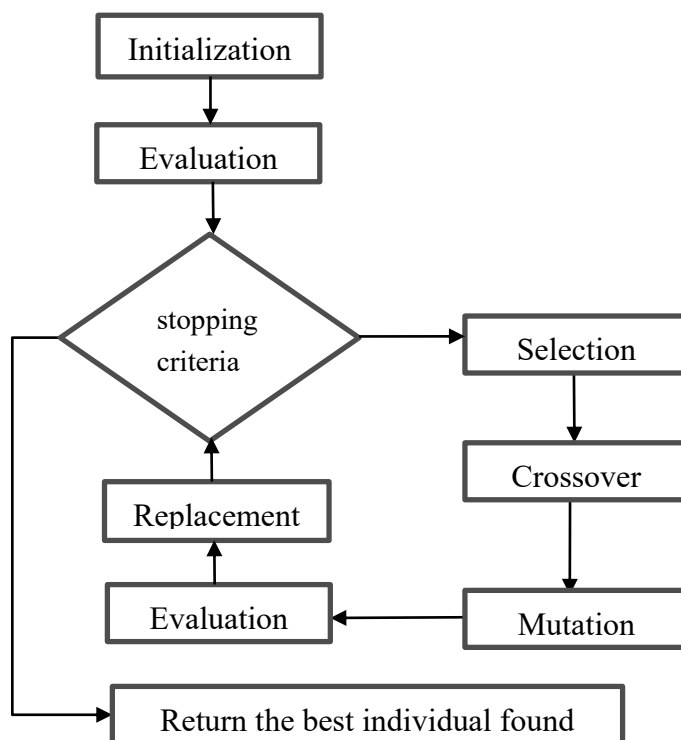
Figure 1: Flow Chart depicting the process of Genetic Algorithm

So the whole figure 1 can be explained as:

    Step 1: randomly initialize population

    Step 2: ascertain fitness of population

    Step 3: repeat

        a.   choose the parents from population

        b.   execute crossover on parents creating population

        c.   execute the mutation of population

        d.   resolve fitness of population

    Step 4: until best individual is good enough

       The above figure 1 depicts the Crossover, the critical genetic operator that permits new solution regions in the search space to be surveyed. It is an arbitrary mechanism for exchanging genes between two chromosomes using the different crossover techniques like - one point crossover, two point crossover, or homologue crossover. In mutation the genes may sometimes be altered, that is in binary code, genes change from either 0 to 1 or vice versa. The offspring always replaces the old population using the elitism or miscellany replacement policy and forms a new population in the next generation [3].

Genetic Algorithms Overview

      Genetic Algorithms [4] are borne on an analogy with the genetic structure and behavior of chromosomes. They are specified within a population of individuals by means of the following foundations:

- Individuals in a population contend for resources and mates.
- Those individuals who are most successful in each 'competition' will produce more descendants than those individuals that accomplish poorly.
- Genes from `good' individuals propagate throughout the population so that two good parents will sometimes produce offspring who are much better than either parent.
- Thus each successive generation will become more appropriate to their environment.

LITERATURE REVIEWS ON GENETIC ALGORITHM

Zhanchao Li et al., [5], conducted an experiment on protein science. A prior information about function of G protein-coupled receptors (GPCRs) can offer beneficial information to pharmaceutical research. The purpose of their function is quite significant theme in protein science. However, with the speedy increase of GPCRs structures entering into databanks, the breach between the amount of known sequence and the amount of known function is widening quickly. Both are time taking and costly to define their function based only on investigational practices. Hence, it is vitally significant to develop a computational process for swift and precise classification of GPCRs.This study brought in a novel three-layer predictor based on support vector machine (SVM) with feature selection, which is established for forecasting and categorizing GPCRs directly from amino acid classification data. The maximum relevance minimum redundancy (mRMR) is exerted to pre-assessed features with discriminative data while genetic algorithm (GA) is used to find the optimized feature subcategories. SVM is utilized for the production of classification models. The complete accurateness with three-layer predictor at levels of superfamily, family and subfamily are obtained by cross-validation assessment on two non-redundant datasets. The results are 0.5% to 16% higher than those of GPCR-CA and GPCRPred. The results with high achievement rates specify that the anticipated predictor is a beneficial automated device in forecasting GPCRs. GPCR-SVMFS, is a corresponding implementable program for GPCRs forecast and sorting that can be acquired freely on appeal from the researchers.

Shuangyin Liu et al., [6] worked on prediction model based on support vector regression (SVR) to solve the aquaculture water quality prediction problem. The SVR parameters must be set carefully to build an effective SVR model. A hybrid approach known as real-value genetic algorithm support vector regression (RGA–SVR) study was presented, which searches using real-value genetic algorithms for the optimal SVR parameters and adopts the optimal parameters to construct the SVR models. To predict the aquaculture water quality data collected from the aquatic factories of YiXing in China, this approach is applied. From actual experiments using monitored aquaculture water quality data, the hybrid approach of SVR with genetic algorithm optimization is able to provide reliable data on the water quality prediction of large-scale intensive aquaculture. Based on the root mean square error (RMSE) and mean absolute percentage error (MAPE), the test results demonstrate that RGA–SVR outperforms the traditional SVR, back-propagation neural network models. To predict aquaculture water quality, the RGA–SVR model is proven to be an effective approach.

SU Bai-hua, and WANG Ying-lin   [7] considered that in natural language processing, Semantic textual similarity is a common task. This measures the degree of semantic equivalence of two textual snippets. In recent times, machine learning methods have been put on to this task,

including methods based on support vector regression. Still, there happens to be number of features consisting of the learning process. Certain features are noisy and irrelative to the outcome. In addition, the prediction performance of the SVR model will be significantly affected by different parameters. The researchers recommend genetic algorithm to select the effective features and optimize the parameters in the learning process all together. To assess the proposed method, they adopted the STS-2012 dataset in the trial. The anticipated GA-based method has improved regression performance when compared with the grid search.

Jui-Sheng Chou et al., [8] studied the hybrid system as an able device to dole out with construction engineering and management problems. The researchers recommend an optimized hybrid artificial intelligence model to assimilate a fast messy genetic algorithm (fmGA) through a support vector machine (SVM). In the initial phase of public-private partnership projects the fmGA-based SVM also called GASVM is utilized for initial forecast of dispute tendency. Predominantly, the SVM provides learning and curve fitting whereas the fmGA optimizes SVM parameters. To find the performance evaluation of proposed hybrid intelligence classification model area under the curve and synthesis index, the measures are taken in form of accuracy, precision, sensitivity and specificity. When compared to other baseline models (i.e., CART, CHAID, QUEST and C5.0) and previous works, investigational assessments specify that GASVM attains improved cross-fold forecast accurateness. The forecasting results prove the proactive warning and decision-support information desired to accomplish potential disputes.

To screen the features of a dataset GA was chosen by Jack and Nandi ; Shon et al., [9], [10]. The chosen subset of features is then fed into the SVM for classification testing. A GA-based technique to find a subset of features for SVM in machine monitoring was developed by Zhang et al., [11]. Applying a GA approach to amend the RBF width parameter of SVM with feature selection was anticipated by Samanta et al., [12]. However, since these approaches only ponder the RBF width parameter for the SVM, they may slip the optimal parameter setting. A GA- based feature selection and parameters optimization for SVM was presented by Huang and Wang [3]. This GA-based feature selection and parameter optimization was utilized for credit scoring.

A study was conducted by Ilhan Ilhan and Gulay Tezel [13] on SNPs (Single Nucleotide Polymorphisms) that are promising tools for disease-gene association studies. However, such a study is inhibited by the high outflow of genotyping millions of SNPs. Hence, it is required to obtain a suitable subset of SNPs to accurately represent the rest of SNPs. Many methods have been developed for the purpose of selecting an appropriate subset of tag SNPs but all of them only offer low prediction accuracy. The authors developed, a new method is developed as GA–SVM with parameter optimization method that benefits from support vector machine and genetic algorithm to forecast as well as to select tag SNPs, respectively. Moreover, it also uses particle swarm optimization algorithm to optimize C and γ parameters of support vector machine. Experiment was carried out using various datasets and the attained results established that, this method can provide better prediction accuracy in recognizing tag SNPs when compared to other methods at present.

GTagger, a heuristic method using genetic algorithms suggested by Mahdevar et al., [14] used correlation and Shannon entropy to compute fitness function. This method results in low prediction accuracy.

The work done by Shutao Li et al., [15] for selecting high discriminative genes from gene expression data has become a vital research. When a vast number of noisy, redundant genes are filtered, this cant only improve the performance of cancer classification but also reduce the rate of medical diagnoses. In this research, a hybrid Particle Swarm Optimization and Genetic Algorithm method are used for gene selection and Support Vector Machine is implemented as the classifier. The suggested method has been tested on three benchmark gene expression datasets like the data of Leukemia, Colon and breast cancer. Experimental results suggest that the proposed strategy can reduce the dimensionality of the dataset and improve classification accuracy.

Mingyuan Zhao et al., [16] performed a study on Generalization performance of support vector machines with Gaussian kernel which is influenced by its model parameters, i.e. error penalty parameter and the Gaussian kernel parameter. A new area distribution model is proposed after researching on the characteristics and properties of the parameter which results simultaneously with Gaussian kernel. The distribution model comprises of optimal straight line, reference point of area boundary, optimal area, transition area, under fitting area and overfitting area. To improve the classification performance of support vector machines, a genetic algorithm based on change area search a new area search is adopted. Comparison of experiments reveal that test accuracy of the genetic algorithm based on change area search a different is better than the two-linear search method.

A study by Y. Rong et al., [17], study recognized results of surface electromyography (sEMG) recorded under circumstances of a maximum voluntary contraction (MVC) in exhausted states using wavelet packet transform and energy analysis. The sEMG signals were listed in 10 young men from the right upper limb with a handgrip. The corresponding energies of certain occurrences were regulated as feature vectors by sEMG signals that were decayed by wavelet packet transform. To discriminate muscle states, a back-propagation neural network, a support vector machine (SVM) and a genetic algorithm-based SVM (GA-SVM) worked as classifiers. The outcomes display that muscle fatigue, MVC could be recognized by level-4 wavelet packet transform and GA-SVM more precisely than when using other methods. The classification accurate rate reached to 97.3% with seven fold cross-validation. The authors claim that proposed method can be used adequately to reflect the muscle activity.

Youngjoo Lee and Jeongjin Lee [18] performed a study on Support vector machine (SVM) with a binary tree architecture that is widespread as it requires the least number of binary SVM to be trained and tested. To design the tree architecture with optimal binary various efforts have been designed. On the other hand, these approaches generally construct a binary tree by a greedy quest. They consecutively divide the classes into two groups so that they consider only local optimum at each node. Genetic algorithm (GA) has been freshly presented in multiclass SVM for the local partitioning of the binary tree structure. The authors proposes a global optimization method of a binary tree structure using GA to develop the classification precision of multiclass problem for SVM. Unlike earlier researches on multiclass SVM using binary tree structures, this present approach universally finds the optimal binary tree structure. The study proposes an enhanced crossover plan to include the determination method of crossover points and generation method of offspring to preserve the maximum information of apparent tree structure for the efficient utilization of GA. Investigational consequences revealed that the planned method provides higher accuracy than any other competing methods used as benchmark within the stipulating time. The

presentation of the method for small size problems is comparable with other challenging methods while more practical developments of the classification accuracy are acquired for medium and large size problems.

Jian Zhao et al., [19] studied on the various treating parameters as model variables. They stressed on quality of plastic part, its warp, volumetric shrinkage and sink marks are to be minimized. The study is based on on two-stage optimization system where the first stage is an improved efficient global optimization (IEGO) algorithm which is adopted to estimate the nonlinear association between processing parameters and the measures of the part quality. In the second stage, to achieve improved spread of design solutions and better convergence near the true Pareto optimal front, non-dominated sorting-based genetic algorithm II (NSGA-II) was used. A cover of liquid crystal display part is optimized to show results that the Pareto fronts obtained by NSGA-II are distributed uniformly. Hence this algorithm has better convergence and vigor. The pair-wise Pareto frontiers illustrate that there is a significant trade-off amid warpage and volumetric shrinkage, and that there is no noteworthy trade-off between sink marks and volumetric shrinkage and between sink marks and warpage.

Chih-Hung Wu et al., [20] developed a genetic-based SVM (GA-SVM) model that can automatically determine the optimal parameters, C and σ, of SVM with the peak predictive accuracy and generalization ability simultaneously. The authors initiated to optimize the parameters of SVM for predicting bankruptcy by employing a real-valued genetic algorithm (GA). The mentioned GA-SVM model was tested on the prediction of financial crisis in Taiwan. The accuracy was compared with other models in multivariate statistics and artificial intelligence. Experimental results of proposed model indicate a very promising hybrid SVM model for predicting bankruptcy in terms of both predictive accuracy and generalization ability. The results of the GA-SVM can stand as a guide of investment for investors and government.

CONCLUSION

In the operations of selection, crossover, and mutation, the GA unites over successive generations towards the global optimum. This simple operation produces rapid, suitable and robust techniques is largely for the reason that GAs combine direction of an effective and efficient procedure. Since population indirectly contain much more evidence than simply the individual fitness scores, GAs combine the good information concealed in a solution with good information from another result to produce many more outcomes with worthy information inherited from both parents, absolutely preceding towards optimality.

The expertise in the application of Genetic algorithm helps to discover and develop simultaneously a growing amount of theoretical validation, and successful application to the real-world problems. It also reinforces the conclusion that Genetic Algorithm is a powerful, effective optimization technique.

REFERENCES

[1]   D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc, 1989.

[2]   J. H. Holland, Adaptation in Natural and Artificial Systems - An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Ann Arbor: The University of Michigan Press, 1975.

[3]   C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," Expert Syst. Appl., vol. 31, no. 2, pp. 231–240, 2006.

[4]   "Introduction to Genetic Algorithms." [Online]. Available:
https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/hmw/article1.html.

[5]   Z. Li, X. Zhou, Z. Dai, and X. Zou, "Classification of G-protein coupled receptors based on support vector machine with maximum relevance minimum redundancy and genetic algorithm," BMC Bioinformatics, vol. 11, no. 325, pp. 1471–2105, 2010.

[6]   S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, and Y. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction," Math. Comput. Model., vol. 58, no. 3–4, pp. 458–465, 2013.

[7]   S. U. Bai-hua and W. Ying-lin, "Genetic Algorithm Based Feature Selection and Parameter Optimization for Support Vector," J. Shanghai Jiaotong Univ., vol. 20, no. 2, pp. 143–148, 2015.

[8]   J.-S. Chou, M.-Y. Cheng, Y.-W. Wu, and A.-D. Pham, "Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification," Expert Syst. Appl., vol. 41, no. 8, pp. 3955–3964, 2014.

[9]   L. B. JACK and A. K. NANDI, "FAULT DETECTION USING SUPPORT VECTOR MACHINES AND ARTIFICIAL NEURAL NETWORKS, AUGMENTED BY GENETIC ALGORITHMS," Mech. Syst. Signal Process., vol. 16, no. 2–3, pp. 373–390, 2002.

[10]   T. Shon, Y. Kim, C. Lee, and J. Moon, "A machine learning framework for network anomaly detection using SVM and GA A Machine Learning Framework for Network Anomaly Detection using SVM and GA," in Proceedings of the 2005 IEEE Workshop on Information Assurance and Security United States Military Academy, 2017, no. February.

[11]   L. Zhang, L. B. Jack, and A. K. Nandi, "Fault detection using genetic programming," Mech. Syst. Signal Process., vol. 19, no. 2, pp. 271–289, 2005.

[12]   B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," Eng. Appl. Artif. Intell., vol. 16, no. 7–8, pp. 657–665, 2003.

[13]   I. Ilhan and G. Tezel, "A genetic algorithm – support vector machine method with parameter optimization for selecting the tag SNPs," J. Biomed. Inform., vol. 46, pp. 328–340, 2013.

[14]   G. Mahdevar, J. Zahiri, M. Sadeghi, A. Nowzari-dalini, and H. Ahrabian, "Tag SNP selection via a genetic algorithm," J. Biomed. Inform., vol. 43, no. 5, pp. 800–804, 2010.

[15]   S. Li, X. Wu, and M. Tan, "Gene selection using hybrid particle swarm optimization and genetic algorithm," Soft Comput., vol. 12, no. 11, pp. 1039–1048, 2008.

[16]   M. Zhao, R. Jian, J. Luping, F. Chong, L. Jianping, and Z. Mingtian, "Parameter selection of support vector machines and genetic algorithm based on change area search," Neural

Comput. Appl., vol. 21, no. 1, pp. 1–8, 2012.

[17]   Y. Rong, D. Hao, X. Han, Y. Zhang, J. Zhang, and Y. Zeng, "Classification of Surface EMGs Using Wavelet Packet Energy Analysis and a Genetic Algorithm-Based Support Vector Machine," Neurophysiology, vol. 45, no. 1, 2013.

[18]   Y. Lee and J. Lee, "Expert Systems with Applications Binary tree optimization using genetic algorithm for multiclass support vector machine," Expert Syst. Appl., vol. 42, pp. 3843–3845, 2015.

[19]   J. Zhao, G. Cheng, S. Ruan, and Z. Li, "Multi-objective optimization design of injection molding process parameters based on the improved efficient global optimization algorithm and non-dominated sorting-based genetic algorithm," Int. J. Adv. Manuf. Technol., vol. 78, no. 9, pp. 1813–1826, 2015.

[20]   C. H. Wu, G. H. Tzeng, Y. J. Goo, and W. C. Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," Expert Syst. Appl., vol. 32, no. 2, pp. 397–408, 2007.