Genetic Algorithms in Pattern Recognition: A Review

1Parul Saxena, 2Dr. Ashish Mehta
1Assistant Professor and Campus Head, Department of Computer Science, Kumaun University, S.S.J. Campus, Almora, Uttarakhand, India
2Assistant Professor and Head, Department of Computer Science, Kumaun University, D.S.B. Campus, Nainital. Uttarakhand, India

Abstract: In the present work, we have studied the basic concepts of pattern recognition, and genetic algorithms then we have made an analysis of application areas of genetic algorithms in various streams of pattern recognition, in which finger print matching, face recognition, optical character recognition, optical feature recognition and disease diagnostic systems are the most important. We have also discussed the basic methodologies used in the optimality of GA and feature selection of objects. At last we have made the conclusion that genetic algorithms can be used very efficiently to converge the results in pattern recognition and in future it can be used in some unexplored areas as efficient leaf recognition in plants and trees.

Index Terms: Genetic Algorithms, Pattern Recognition, Pattern Recognition Models.

I. INTRODUCTION

Pattern Recognition is very important branch of computer science, nearly synonymous with machine learning. This branch of artificial intelligence focuses on the recognition of patterns and regularities in data. In many cases, these patterns are learned from labeled training data in case of supervised learning, but when no labeled data is available other algorithms can be used to discover previously unknown patterns as in unsupervised learning. The terms pattern recognition, machine learning, data mining and knowledge discovery in databases are hard to separate, as they largely overlap in their scope. Machine learning is the common term for supervised learning methods and originates from artificial intelligence. Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform most likely matching of the inputs, taking into account their statistical variation. This is opposed to pattern matching algorithms, which look for exact matches in the input with pre-existing patterns [41][43].

The Genetic Algorithm is a model of machine learning which derives its behavior from a metaphor of the processes of evolution in nature. This is done by the creation within a machine of a population of individuals represented by Chromosomes, in essence a set of character strings that are analogous to the base-4 chromosomes that we see in our own DNA. The individuals in the population then go through the process of evolution [1][2][3][4]. In nature, we see that the encoding for our genetic information (GENOME) is done in a way that admits a sexual reproduction such as by budding typically results in offspring that are genetically identical to the parent. Sexual reproduction allows the creation of genetically radically different offspring that are still of the same general flavor. The Crossover operation happens in an environment where the selection of who gets to mate is a function of the fitness of the individual, i.e. how good the individual is at competing in its environment. Some Genetic Algorithms use a simple function of the fitness measure to select individuals to undergo genetic operations such as crossover or a sexual reproduction [19][20].

II. GENETIC ALGORITHMS IN PATTERN RECOGNITION

Here, the various contributions of GA in pattern identification and classification have been analyzed as follows:

A. K-Nearest Neighbour with GA

K-Nearest Neighbour (KNN) is one of the most popular algorithms for pattern recognition. Many researchers have found that the KNN algorithm accomplishes very good performance in their experiments on different data sets. The traditional KNN text classification algorithm has three limitations:

(i) Calculation complexity due to the usage of all the training samples for classification,
(ii) The performance is solely dependent on the training set, and
(iii) There is no weight difference between samples.

To overcome these limitations, an improved version of KNN is discussed in [27] and Genetic Algorithm (GA) is combined with KNN to improve its classification performance. Instead of considering all the training samples and taking k-Neighbours, the GA is employed to take k-Neighbours straightaway and then calculate the distance
to classify the test samples. Before classification, initially the reduced feature set is received from a novel method based on Rough set theory hybrid with Bee Colony Optimization (BCO).

B. GA in Point Pattern Matching
Point pattern matching (PPM) is an important topic in the fields of computer vision and pattern recognition. If there exists a one to one mapping between the two point sets to be matched, PPM can be divided into the case of complete matching and the case of incomplete matching. PPM can be divided into labelled point-matching case and unlabelled point-matching case. Using partial Hausdorff [28] presented a genetic algorithm (GA) based method to solve the incomplete unlabelled matching problem under general affine transformation. Since it successfully reduces the solution space of GA by constructing feature ellipses of point sets, the method can achieve high computing efficiency and good matching results.

C. Optical Feature Representation with GA
An approach to online handwritten signature verification based on optimal feature representation (OFR) and neural-network-driven fuzzy reasoning (NND-FR) is discussed in [29]; OFR is different from feature selection and generation. To create a reference signing model of a person, a set of shape features and dynamic features are extracted from a set of original signatures.

D. GA in OCR to Optimize Classification Error
The methods for Optical Character Recognition (OCR) system using genetic algorithm are investigated in [30]. Genetic algorithm is an unbiased optimization algorithm which makes parameter selection in an optimized way so as to obtain the global optimum. The genetic algorithm is used for optimizing the classification error.

E. Hybrid Algorithm for Artificial Immune System
The study of artificial immune systems is a relatively new field that tries to exploit the mechanisms of the natural immune system (NIS) in order to develop problem solving techniques. Genetic Hybrid algorithm for artificial immune system in [31] modified the clonal selection algorithm, which is inspired from the clonal selection principle and affinity maturation of the human immune responses, by hybridizing it with the crossover operator, which is imported from GAs to increase the exploration of the search space. The adaptability of the mutation rates by applying a degrading function has been introduced so that the mutation rates decrease with time where the affinity of the population increases, the hybrid algorithm used for evolving a fuzzy rule system to solve the well known Wisconsin Breast Cancer Diagnosis problem (WBCD).

F. Distributed Hierarchical Genetic Algorithm (DHGA)
Distributed hierarchical genetic algorithm (DHGA) for optimization and pattern matching has been introduced in [32]. It is eventuallly a hybrid technique combining the advantages of both distributed and hierarchical processes in exploring the search space. The search is initially distributed over the space and then in each subspace the algorithm works in a hierarchical way. The entire space is essentially partitioned into a number of subspaces depending on the dimensionality of the space. This is done in order to spread the search process more evenly over the whole space. In each subspace the genetic algorithm is employed for searching and the search process advances from one hypercube to a neighbouring hypercube hierarchically depending on the convergence status of the population and the solution obtained so far.

G. GA and Disease Diagnostic System
Three disease diagnosis systems using pattern recognition based on genetic algorithm (GA) and neural networks have been discussed in [33]. All systems dealt with feature selection and classification. GA chose subsets of features for the input of the classifier (neural network) and the accuracy of the classifier determined the percentage of effectiveness of each subset of features. The classifiers using in [33] were general regression neural network (GRNN), radial basis function (RBF) and radial basis network exact fit (RBEF). It uses breast cancer and hepatitis disease datasets taken from UCI machine learning database as medical dataset. The system performances were estimated by classification accuracy and they were compared with similar methods without feature selection.

H. GA and Fitness Control
The fitness control procedure [34] for the application of genetic algorithms (GA) in pattern recognition is very important. Instead of using GA to solve an optimization problem, the task is rather represented as a set of data bit strings, and a similarity measure among those data bit strings. Then, the GA is used to select the subgroup yielding highest mutual similarity.

I. GA in Character Recognition
Character Recognition is the mechanical or electronic translation of scanned images of handwritten, typewritten or printed text into machine-encoded text. In India, more than 300 million people use Devanagari script for documentation. There has been a significant improvement in the research related to the recognition of printed as well as handwritten Devanagari text in the past few years. The problem arises in Devnagari script character recognition using quadratic classifier provides less correctness and less efficiency. For the answer of the above problem and for getting better efficiency genetic algorithm method is used in [35]. The idea of genetic algorithm comes from the fact that it can be used as an outstanding means of combining various styles of writing a
character and generates new styles. Closely observing the ability of human mind in the recognition of handwriting, we find that humans are able to recognize characters even though they might be seeing that style for the first time. This is possible because of their power to visualize parts of the known styles into the unknown character.

J. Human as Information Processing System
A human being is a very sophisticated information processing system (IPS) [36][52], partly because he possesses a superior pattern recognition capability. Though human beings have very good pattern recognition ability, human sensory system has certain limitations. Some undetectable patterns or otherwise patterns which are more complex in nature become difficult for human beings to recognize. The tedious tasks require huge manpower. Information can be the most valuable asset to the human being only if he can extract potential or valuable knowledge from it. Human analysts can no longer keep track of or make sense of the enormous volume of information in this era of rapidly growing and advanced technology. Thus, it becomes essential to automatically process this excess of information efficiently and automatically.

K. Pattern Recognition System
A typical Pattern Recognition system (PRS) [36],[50],[51], [53] consists of three tasks namely, data acquisition and/or preprocessing, feature analysis, classification and clustering. In the first step, data are collected by using some sensors or other means and then these raw data may be preprocessed. Preprocessing may involve noise reduction, normalization and conversion of raw data into suitable form for pattern recognition. After obtaining the data, good features are extracted by mapping data to other domain or a subset of good features is selected from the available features.

L. Evolutionary Algorithms in Biometric Systems
The wide usage of biometric information for person identity verification purposes, terrorist acts prevention measures and authentication process simplification in computer systems has raised significant attention to reliability and efficiency of biometric systems. Modern biometric systems still face many reliability and efficiency related issues such as reference database search speed, errors while recognizing of biometric information or automating biometric feature extraction. Current scientific investigations show that application of evolutionary algorithms may significantly improve biometric systems [37][38][39].

M. GA for Surgically Altered Face Images Recognition
A multi-objective evolutionary granular algorithm to match face images before and after plastic surgery has been discussed in [40]. The algorithm first generates non-disjoint face granules at multiple levels of granularity. The granular information is assimilated using a multi objective genetic approach that simultaneously optimizes the selection of feature extractor for each face granule along with the weights of individual granules. On the plastic surgery face database, this algorithm yields high identification accuracy as compared to existing algorithms and a commercial face recognition system.

N. GA and SGA in Face Recognition
Face Recognition is a progressive field of research which is facing many challenges. A method of genetic algorithm (GA) based on neural network for feature selection that retains sufficient information for classification purposes is presented in[41][42][43][44]. This method presents an integration of genetic algorithm with an artificial neural network classifier.

The concept of spiral genetic algorithm (SGA) is incorporated which decreases the search area of the subsequent GA’s that is proportional to the minimal cost function of the previous GA and thus decreasing the cost to provide better convergence [45].

III. METHODS
Recognition is a well-developed area in the field of pattern recognition. Here, as we introduce the basic concepts and algorithm developments in pattern recognition through GA.

A. Outline of the Genetic Algorithm
The following outline summarizes how the genetic algorithm works [16][17][18]:
1. The algorithm begins by creating a random initial population.
2. The algorithm then creates a sequence of new populations. At each step, the algorithm uses the individuals in the current generation to create the next population. To create the new population, the algorithm performs the following steps:
a) Scores each member of the current population by computing its fitness value.
b) Scales the raw fitness scores to convert them into a more usable range of values.
c) Selects members, called parents, based on their fitness.
d) Some of the individuals in the current population that have lower fitness are chosen as elite. These elite individuals are passed to the next population.
e) Produces children from the parents. Children are produced either by making random changes to a single parent-mutation or by combining the vector entries of a pair of parents-crossover.
f) Replaces the current population with the children to form the next generation.
3. The algorithm stops when one of the stopping criteria is met [5][6][7][8].
The unary and higher order transformations are called evolutionary operators. The two most frequently evolutionary operators are:

**B. Mutation**

It modifies an individual by a small random change to generate a new individual [10][11]. This change can be done by inverting the value of a binary digit in the case of binary representations, or by adding (or subtracting) a small number to (or from) selected values in the case of floating point representations. The main objective of mutation is to add some diversity by introducing more genetic material into the population in order to avoid being trapped in a local optimum. Generally, mutation is applied using a low probability. However, some problems (e.g. problems using floating point representations) require using mutation with high probability [9].

**C. Recombination (or Crossover)**

In this phenomenon parts from two (or more) individuals are combined together to generate new individuals [11]. The main objective of crossover is to explore new areas in the search space [9].

**D. Fitness Function**

A key element in GAs is the selection of a fitness function that accurately quantifies the quality of candidate solutions. A good fitness function enables the chromosomes to effectively solve a specific problem. Both the fitness function and solution representation are problem dependent parameters. A poor selection of these two parameters will drastically affect the performance of GAs. One problem related to fitness functions that may occur when GAs are used to optimize combinatorial problems is the existence of points in the search space that do not map to feasible solutions. One solution to this problem is the addition of a penalty function term to the original fitness function so that chromosomes representing infeasible solutions will have a low fitness score, and as such, will disappear from the population [23].

**E. Selection**

Another key element of GAs is the selection operator which is used to select chromosomes (called parents) for mating in order to generate new chromosomes (called offspring). In addition, the selection operator can be used to select elitist individuals. The selection process is usually biased toward fitter chromosomes. Selection methods are used as mechanisms to focus the search on apparently more profitable regions in the search space [24][25][26].

**F. Stopping Conditions for the Algorithm**

The genetic algorithm uses the following conditions to determine when to stop:

- **Generations**: The algorithm stops when the number of generations reaches the value of Generations.
- **Time limit**: The algorithm stops after running for an amount of time in seconds equal to Time limit.
- **Fitness limit**: The algorithm stops when the value of the fitness function for the best point in the current population is less than or equal to Fitness limit.
- **Function Tolerance**: The algorithm runs until the average relative change in the fitness function value over Stall generations is less than Function tolerance.
- **Nonlinear constraint tolerance**: The Nonlinear constraint tolerance is not used as stopping criterion. It is used to determine the feasibility with respect to nonlinear constraints. Also, a point is feasible with respect to linear constraints when the constraint violation is below the square root of nonlinear constraint tolerance [21][22]

The algorithm stops as soon as any one of these conditions is met. We can specify the values of these criteria in

**G. PR as optimization problem**

Most pattern recognition tasks can be viewed as optimization problems. There are many tools available that can be used for optimization. The suitability of the techniques depends on the objective function, the constraints and the control variables. If the objective and the constraints are linear functions of the variables, then such a problem is called a linear programming problem and it can be solved using methods like Simplex Method. Calculus-based optimization methods are suitable for problems having smooth, continuous, and differentiable (search) surfaces. These methods use gradient information to find the optimal solution. Hill climbing is a popular calculus-based method. It climbs the surface in the steepest permissible direction to find local optima. These methods fail to give global optima which are desired. Unfortunately, many real world problems are complex having discontinuous, multi modal surfaces without existence of derivatives. If the search space is small, then enumerate search techniques can be used to find the optimal solution. In this approach, the objective function at every point of the search space is evaluated to determine the optimal one. This is only suitable for very small (search space) problems. Even for moderate size, the computational cost is so high that it is not practically possible to be implemented. To reduce computational time, we may implement techniques like "branch and bound" and "Dynamic Programming". These repeatedly partition the problem into a set of smaller sub-problems to reduce search time. But these too fail when search space increases. Consider popular traveling salesman problem (TSP). In this problem, we minimize the distance taken by a traveling salesman who has to visit a certain number of cities exactly once and return home. When the number of cities is very small, then the number of possible solutions is relatively small and hence we can easily find the optimal solution using exhaustive search [36]. For example, with 5 cities, all possible routes can be easily checked. However, for a 50-
city problem, the number of solutions rises to 10^60. This is very vast search space where it is almost impossible to find the optimal solution [12][13][14][15].

IV. CONCLUSION

In this work, first we have studied about Pattern Recognition and Genetic Algorithms and then we have analyzed the GA based methodologies to solve certain pattern recognition tasks such as finger print matching, face recognition, optical character recognition, optical feature recognition etc. Also we have analysed the various aspects of Genetic algorithms as optimality and Particle Swarm Optimization and its various applications in Pattern Recognition [46][47][48][49]. During this study we observe that the genetic algorithms can be used very efficiently to optimize the results to solve the real world pattern recognition problems. In future the unexplored application areas of Genetic algorithms in pattern recognition as measuring the color difference between two objects, or to recognize the palms of various plants, are open for research.

References

[38] Kyre Glette, Jim Torresen, Moritoshis Yasunaga, An Online EHW Pattern Recognition System Applied to Face Image
Recognition, project report (no. 160308/V30) in Biological-Inspired Design of Systems for Complex Real-World Applications.


