

Modified Genetic Algorithm Parameters to Improve Online Character Recognition Accuracy

Abstract

Online character recognition is characterized with feature extraction and classification parameters that make recognition accuracy non-trivial task. Failure of existing optimization techniques to yield an acceptable solution to solve poor feature selection and slow convergence time provokes the idea for some stochastic algorithms. In this paper, a feature reduction technique that apply the power of genetic algorithm was modified using fitness function and genetic operators to minimize the aforementioned drawbacks. Two classifiers (C1 and C2) were then formulated from the integration of modified genetic algorithm(MGA) into an existing Modified Optical Backpropagation (MOBP) learning algorithm. The performance of C2 on generation gaps and database sizes were further evaluated using convergence time and recognition accuracy. The research evaluation showed that C2 assumed average convergence times of 130.30, 211.69, 199.23 and 243.00 milliseconds with generation gaps of 0.1, 0.3, 0.5 and 0.7. This implies that generation gap variation had a positive effect on the network performance..

Keywords

character recognition, genetic algorithm, feature extraction, feature selection, genetic operators and generation gap.

1. Introduction

Genetic algorithm(GA) is one of the widely used biological evolution technique for global optimization that mimic the principle of natural genetics and natural selection. GA is an iterative algorithm based on many generations of probable solutions, among selection schemes authorization the removal of bad solutions and the reproduction of good ones that can be modified[1][2]. Character recognition(CR) refers to the mechanical or electronic conversion of images of handwritten or printed text into machine-editable text. However, most existing classifiers used in recognizing online handwritten characters suffer from dynamic of poor feature selection and slow convergence which affect training time and recognition accuracy[3][4]. Online character recognition constitutes a serious challenges: the complexity of noise from data, variations in character styles and mood of the writers make it difficult[5]. Adequate modeling of online character recognition mostly depends on the choice of a certain feature selection and classification procedures. Some conventional optimization methods cannot overcome these limitation, genetic algorithm is introduced as a good alternative optimization methods for solving such problems [6]. GAs popularity, like its global perspective, wide spread applicability, inherent parallelism and noise tolerance make its suitable in various search and optimization problems and character recognition accuracy[7]. However, the effectiveness of a certain optimization technique can be evaluated by the recognition accuracy achieved and the convergence time needed. Feature selection algorithm is to reduce the number of features extracted to improve classification task performance. Different feature selection algorithms are used, which include Principal component analysis, Correlation-based feature selection, Discrete wavelet transform, Linear discriminant analysis, Gabor filters and many others. In this paper MGA was used to find the optimal feature subset. The reasons for performing feature selection as stated by [8][9] was to lower computational time, reducing storage requirements, improving data understanding and better classification performance. The generation gap is the proportion of chromosomes in the population which are replaced in each generation. It is a genetic algorithm parameter used for high performance solution and greedier because it always uses the best individual of population for generating offspring [10]. Hence, the objective of the paper is to determine whether how significant the effect of generation gap as a sensitive genetic algorithm parameter towards recognition accuracy and convergence time. The power of genetic algorithm parameter is used to overcome the drawback of spending much time for character to converge.

52 2. Related Work

53 Previous work on online handwritten character recognition had shown that feature extraction process is
 54 most important factor in achieving high recognition performance. A number of methodologies have been
 55 proposed over the years for character recognition and some of them have outstanding performance. [11]
 56 developed a genetic based neural network system for online character recognition. This research
 57 integrated modified genetic algorithm into modified backpropagation neural network to improve the
 58 performance of character recognition system. This paper investigated the necessity for optimization
 59 algorithms to enhance the performance of three classifiers. The recognition accuracy obtained for
 60 developed model shown a better accuracy than other classifiers. [12] developed a paper on, conversion of
 61 handwritten data into electronic data, using neural network approach and Genetic Algorithm to recognize
 62 handwritten characters. The handwritten characters were converted to graphs and was intermixed to
 63 generate new or unique styles intermediate between the styles of parent character. [13] was the first to
 64 evaluate empirically the performance of GAs with overlapping populations. He introduced the generation
 65 gap G as a parameter to the GA, and found that at low values of G the algorithm had a severe loss of alleles,
 66 which resulted in poor search performance. [14] developed English character recognition system using
 67 the hybrid of standard backpropagation and genetic algorithm for recognition of uppercase alphabets.
 68 Recognition performance was 91.1%. Furthermore, [15] presented a backpropagation network algorithm
 69 combined with genetic algorithm to achieve both accuracy and training swiftness for recognizing
 70 alphabets. [16] presented a methodology to recognize handwritten character written on the digitizing
 71 tablet. It uses different spatial and temporal features to extract features from the character and genetic
 72 algorithm was used as a tool to find an optimal subset. The overall experimental results in recognition
 73 rate was 83.1%. The main goal is to further improve the recognition accuracy and convergence time using
 74 modified genetic algorithm and MOBP at classification level.

75 3. Material and Methods

76 3.1 Database Acquisition and Normalization

77 Fig. 1 shows the block diagram of the character recognition model which consists of character acquisition,
 78 preprocessing, feature extraction and selection, and classification. The dataset was locally acquired using
 79 a pen digitizer tablet contains 50 randomly selected students of Yaba College of Technology, Yaba.
 80 Nigeria. Character information of 26 upper case (A-Z), 26 lower case (a-z) English alphabets and 10 digits
 81 (0-9) were considered making a total of 62. The students were asked to write for each character twice to
 82 allow the network learn various possible variations the characters and become adaptive in nature. All
 83 these entries are stored and served as the training dataset which was the input data that was fed into
 84 online character recognition system. The raw data collected were subjected to a number of preprocessing
 85 steps(i.e. binarization, normalization, extreme coordinate measurement and grid resizing). This is to
 86 convert the character image to binary character, removal of noise and redundancy and the output was
 87 normalized by resizing to a matrix size of 5 by 7 for feature extraction

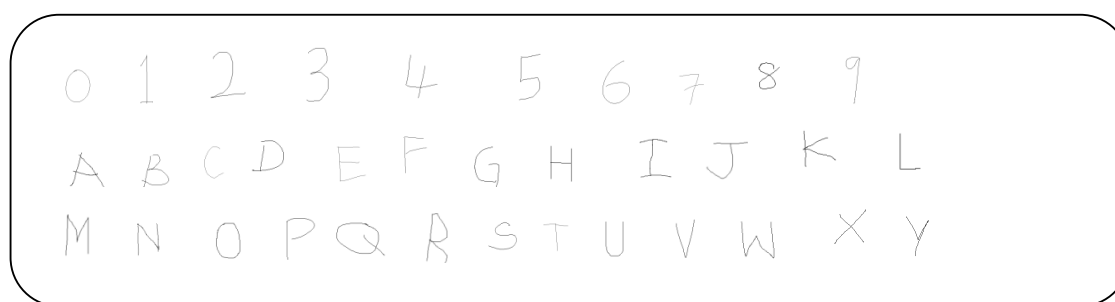


Figure 2 shows a sample of dataset used

105 3.2 Experimentation of the Model

The process of features extraction were employed to extract the features from character to aid classification task. The extracted features consist of stroke information, invariant moments, projection and zoning of the handwritten character to create a global feature vector. The output of the extracted and selected features of the handwritten character was fed into the digitization stage in order to convert the extracted features into digital form. The paper combines structural and statistical features to complement each other, handle style variations and highlight properties that identity a character. Two unique strategies were modified to improve the standard GA optimization speed; mutation is used to prevent falling solution into local minima and preventing the chromosomes from becoming too similar to each other. Secondly, modification of fitness function to minimize the recognition errors.

$$\text{Fitness} = (\text{no of selected zero features}) * \% \text{FeatureUsed} \quad (1)$$

where $\text{FeatureUsed} = (\text{Hprojection} / (\text{Downsample_Height} * \text{Downsample_Width}) * 100)$.

Then followed by feature selection using MGA. The aim is to retain only discriminant features and choose ;optimal set from original features for classification task. This minimizes the computational burden imposed by using many features, reduce weighty matrix and character patterns.

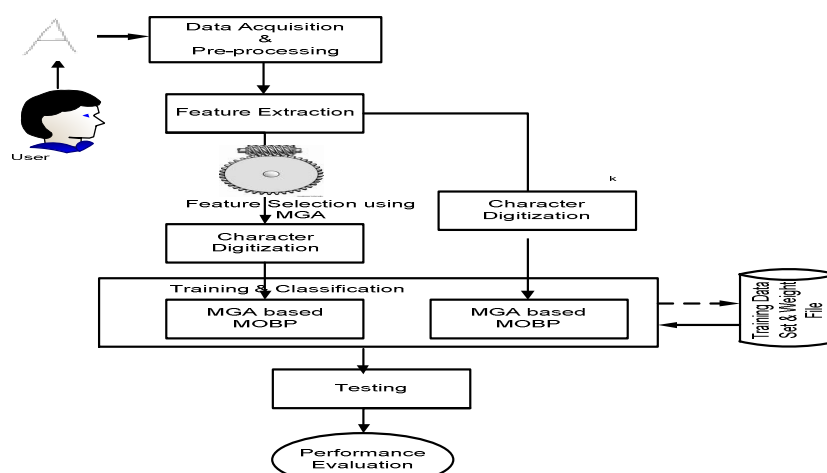


Figure 1: Block diagram of the Developed Online Handwritten Character Recognition Model

The extracted features were subjected to optimization technique (MGA) for optimal feature selection. The initial inputs of the characters provided are trained with different parameters and different sets of characters. The block diagram shows the training, classification and testing and the training phase contains the handwritten database that was trained. The Model hybridized the modified genetic algorithm and modified optical backpropagation neural networks for the training and classification of the input pattern. The initial inputs of the handwritten characters are trained with different parameters and sets of characters. The characters are classified using three classifiers: C1 and C2. During training, the weights of the network are iteratively adjusted to minimize the error function. The set of outputs obtained are fed into the genetic algorithm to select the best fittest and best solution. The outputs of the genetic algorithm are sent to the neural networks as input. The recognition model involves the usage of artificial neural networks which have been exhaustively trained to recognize different types of handwritten This is achieved by using highly efficient supervised learning algorithm. ;The output obtained from the trained are stored as files. Match the introduced character with the one in the database template and classify the given character or pattern image. The genetic algorithm is used for optimal feature selection. The extracted features are classified according to similarity of their shapes and features from the dataset collected from different individuals.

3.2.1 The Genetic and Neural Network Training Algorithm

The training algorithm involves the following two stages:

Stage A: Performs the training of the character sample generated using Genetic Algorithm.

Step 1: Select character samples collected from various people (Generate random population of n chromosomes this corresponds to initial population for the genetic Algorithm).

Step 2: Evaluate the fitness function of the population equation

Step 3: Mutate new offspring at each locus based on mutation probability.

Step 4: Select parent chromosomes from the population according to their fitness function .

Step 5: Perform crossover on two strongest parents to create a new offspring and replace the weakest parent with the new child created.

Step 6: Insert the new offspring in the new population and use new generated population for a further run of the algorithm.

STAGE B: Performs the training of the input from the genetic algorithm to the output layer. With the introduction of cubic error function, the modified Optical Backpropagation is given as:

1. Input vector, $X_p = X_{p1} \mid X_{p2} \mid X_{p3} \dots \mid X_{pm}$ to the input units. Each value in input layer receive X_{pi} and forward to values in hidden layer.

2. For each character, calculate the output of the feed forward network

3. Calculate the cubic error term by comparing with the desired output corresponding to the symbol

Using cubic error adjustment where

Y_{pk} = Desired output, O_{pk} = Network output

$$T = (Y_{pk} - O_{pk})^2$$

4. Back propagate error across each link to adjust the weights

5. Move to the next character and repeat steps 2 to 4 until all characters were visited

6. Compute the average error of all characters

7. Repeat steps from steps 2 - 6 until the cubic error was acceptably small or error threshold is reached for each of the training vector pair.

3.2.2 Classification and Testing Stage

The classification phase is the last stage of the developed character recognition system. This phase determined the overall performance of the developed algorithm. Two types of classifiers were implemented for online character recognition.

(a) C1 (MGABased MOBP)

Two classifiers (C1 and C2) were formulated from MGA-MOBP such that C1 classified using MGA at classification level. There was no optimization at feature selection, therefore all the features were used. The MGA was used at classification level.

(b) C2 ((MGA and MGABased MOBP)

C2 employed MGA at feature selection level and classified at classification level. The performance of each of C1 and C2 was evaluated using recognition failure, training time and correct recognition accuracy. The objective for using the optimization algorithm at feature selection level is in twofold: (i) feature reduction which improves the convergence time and (ii) elimination of insignificant features which improves the classification accuracy and reduce computation time imposed of whole features. In this paper, a modified genetic algorithm was used for optimal feature selection. MGA was used during the training to solve the problem of local minima in backpropagation and find the optimal set weights at a very short time. The recognition of character is done based on minimum distance classifier between two feature vectors (testing and template of training vectors) which forms the character input pattern using Euclidean distance metric. That is template of When the minimum distance is less or equal to a threshold set implies correct match or a recognized character.

4 Experimental Design and Results

4.1 Experimental Setup

The development tool used is Microsoft Visual C# version 4.0 on Windows 8 Ultimate 64-bit operating system, Intel@CoreTM i5-3210M CPU @2.50 GHz processor, 8GB Random Access Memory. The decision taken to recognize or classify the images was based on training time, recognition time and recognition accuracy. In this study, training and test sets consist of 6200 and 540 sample images respectively.

4.2 Evaluation Results

Table 1 represented the results of analysis of influence of generation gap on average recognition accuracy and average convergence time. This was achieved by varying the generation gap through 0.1, 0.3, 0.5, and 0.7. The best convergence time and recognition accuracy was obtained as Ggap 0.1 which is 103.20 milliseconds and 99.44% respectively at 6200 datasets. This implies that generation gap variation had a positive effect on the network performance. At 0.1 Ggap, the average recognition accuracy is proportion inverse to the convergence time.

Table 1: Effect of variation of Generation gap on Average recognition accuracy and Average convergence time.

Generation gap	Average Recognition Accuracy (%)	Average Convergence Time (milliseconds)
0.1	99.44	130.30
0.3	98.83	211.69
0.5	98.51	199.23
0.7	98.11	243.00

Table 2 the recognition accuracy at 0.7 is 97.81% and 99.44% at Ggap 0.1 respectively. The recognition accuracy increases with decrease in Ggap of number . This was due to the heuristic nature of the genetic algorithm and larger character samples in the vector space. The smaller the portion of the population that is replaced each generation the better the average recognition accuracy and convergence time. at the two classifiers, the fitness function values shown little or no significant level.

Table 2: Performance evaluation of the classification accuracies of two classifiers

Ggap	C1				C2			
	Fitness function	CR (%)	FR (%)	RF (%)	Fitness function	CR (%)	FR (%)	RF (%)
0.7	0.245	96.7	3.0	0.3	0.245	98.1	2.1	0.1
0.5	0.245	97.4	2.6	0.0	0.245	98.5	1.3	0.2
0.3	0.243	98.3	1.7	0.0	0.244	98.8	1.2	0.0
0.1	0.245	98.7	1.3	0.0	0.246	99.4	0.6	0.0

CR = Correct Recognition, FR = False Recognition and RF =Recognition Failure

4. Conclusion and Future work

In conclusion, MGA was used for feature selection to reduce the feature space which enhanced recognition accuracy and reduced training time; therefore the primary objective of the study is achieved. A model for online character recognition was proposed using MGA and MOPB algorithm for two classifiers: C1 and C2 for overall recognition accuracy. In this investigation altogether MGA has been examined. The influence of genetic algorithm parameter, namely, generation gap, has been employed for aiming to improve the recognition accuracy and convergence time. The results show that there is better accuracy and convergence time as generation gap values increases. The results obtained show that the convergence time using generation gap of 0.1 instead of 0.7 was the best and reliable. As a result of the findings during the course of study, it is recommended that there could be further research on word and sentence recognition systems with MGA & MOBP as classifier.

References

1. Kherallah, M., et al. (2009): "On-line Arabic Handwriting Recognition System Based on Visual Encoding and Genetic Algorithm", Engineering Applications of Artificial Intelligence, 22(1):153-170.
2. Noaman KMG, Saif JAM, Alqubati IAA. Optical character recognition based on genetic algorithms. Journal of Emerging Trends in Computing and Information Sciences. 2012;6(4):203-208.
3. Adigun, JO., Omidiora, EO., Olabiyisi, SO., Fenwa, OD. Oladipo, O., and Rufa, MM. (2015)" Development of a Genetic based Neural Network System for Online Character Recognition", International Journal of Applied Information Systems (IJ AIS), 9(3):1-8.

2384. Adeyanju, IA, Ojo, OS. and Omidiora, EO." Recognition of Typewritten Characters Using Hidden Markov
239 Models", , British Journal of Mathematics & Computer Science 12(4): 1-9.
2405. Fenwa, OD., Omidiora, EO., Fakolujo, OA., Ajala, FA., Oke, AO. (2012): "A Modified Optical Backpropagation
241 Neural Network Model in Online Handwritten Character Recognition System", International Journal of
242 Computer Application 2 4(2): 190-201
2436. Angelova M, Pencheva T (2011) Tuning genetic algorithm parameters to improve convergence time. Int. J.
244 Chem. Eng., Article ID 646917. 1-7.
2457. Yeremia Hendy, Niko Adrianus Yuwono, Pius Raymond and Widodo Budiharto (2013): "Genetic
246 Algorithm and Neural Network for Optical character recognition" Journal of computer science 9 (11):
247 1435-1442
2488. Adigun, JO, Fenwa, OD, Omidiora, EO. and Olabiyisi, SO. (2016) 'Optimized Features for Genetic Based Neural
249 Network Model for Online Character Recognition", British Journal of Mathematics & Computer Science 14(6): 1-
250 13.
2519. Aziz , Amira. Sayed A., Azar, Ahmad. Taher, Salama, Mostafa A : " Genetic Algorithm with Different Feature
252 Selection Techniques for Anomaly Detectors Generation", Proceedings of the 2013 Federated Conference
253 on Computer Science and Information Systems. 769-774.
25410. Jani Rokkonen and Jouni Lampinen (2005) A comparison of Differential Evolution and Jumanal Shilpa and
255 Holi, Ganga (2013): "On-line Handwritten English Character Recognition Using Genetic Algorithm"
256 International Journal of Computer Trends and Technology (IJCTT). 4(6): 1885-1890.
25711. Adigun, JO., Omidiora, EO., Olabiyisi, SO., Fenwa, OD (2015)" A Modified Genetic based Neural Network Model
258 for Online Character Recognition", International Journal of Applied Information Systems (IJ AIS), 9(5):18-23.
25912. Samta Jain Goyal and Rajeev Goyal(2015): "An Idea to Recognition of handwritten Characters using
260 Genetic Algorithms", An international journal of advanced computer technology, 4 (4): 1686-1689.
26113. De Jong, K. (2006).Evolutionary computation - a unified approach MIT Press
26214. Ranpreet, K. and Singh, B. (2011): "A Hybrid Neural Approach for Character Recognition System",
263 (IJCSIT) International Journal of Computer Science and Information Technologies, 2 (2):721-726.
26415. Padhi, D (2012):"Novel Hybrid Approach for Odia handwritten Character Recognition System",
265 International Jornal of Advanced Research in Computer Science and Software Engineering, 2(5):150-157.
26616. Jumanal Shilpa, Holi Ganga(2013). On-line handwritten English character recognition using genetic
267 algorithm. International Journal of Computer Trends and Technology (IJCTT)4(6):1885-1890.