

Character Recognition Using Neural Network Learned by Artificial Bee Algorithm

Dr. Bassim A. Jumaa¹, Dr. Ayad Ghazi Naser², Maryam Khalaf³

¹ Dept. Of Computer Engineering and Information Technology, University of Technology

Email: basim_alani@yahoo.com

² Dept. Of Computer Engineering and Information Technology, University of Technology

Email: dr_ayad64@yahoo.com

³ Dept. Of Computer Engineering and Information Technology, University of Technology

Email: mareamkalf@gmail.com

Abstract: Character Recognition is the text recognition system that allows hard copies of written or printed text to be rendered into editable, soft copy versions. In this paper, work has been performed to recognize pattern using multilayer perceptron learning by Artificial Bee algorithm (ABC) that simulates the intelligent foraging behavior of a honey bee swarm. Multilayer Perceptron (MLP) trained with the standard back propagation (BP) algorithm normally utilizes computationally intensive training algorithms. One of the crucial problems with the BP algorithm is that it can sometimes yield the networks with suboptimal weights because of the presence of many local optima in the solutions space. The suggested method is to use ABC for learn the Neural Networks, to solve text character recognition problem, by update the Neural Networks weights. A comparison studies are made between ABC and BP methods in NN learning to specify which is better in solving character recognition problem.

Keywords: Neural Network, Artificial Neural Network, Artificial Bee Algorithm, Pattern Recognition, Back Propagation, and Classification field.

الخلاصة: تمييز الرموز هو نظام تمييز النصوص الذي يسمح بتحويل النسخ المكتوبة او المطبوعة الى نصوص قابلة للتعديل نسخة الالكترونية . ففي هذا البحث تم انجاز عملية تمييز الانماط باستخدام الشبكة العصبية متعددة الطبقات المتعلمة بواسطة خوارزمية مستعمرات النحل (ABC) والتي تحاكي السلوك الذكي لسرب النحل. الشبكة العصبية متعددة الطبقات سبق وان عُلمت باستخدام خوارزمية الانتشار التراجعي القياسية (BP). ان واحدة من المشاكل التي تعانيها خوارزمية (BP) هو انتاج اوزان شبه مثالية بسبب الوقوع في بعض النهايات المثالية المحلية من فضاء الحلول. ان الطريقة المقترحة استخدام (ABC) لتعليم الشبكة العصبية، لحل مسألة تمييز الرموز، من خلال تعديل الاوزان. وايضا تم اجراء دراسة مقارنة بين خوارزمية (ABC) و (BP) عند تعليم الشبكة لتحديد الطريقة الافضل عند حل مسألة تمييز الرموز.

1. Introduction

Humans have developed highly sophisticated skills for sensing their environment and taking actions according to what they observe, e.g., recognizing a face, understanding spoken words,

reading handwriting and distinguishing fresh food from its smell. Pattern recognition wants to give similar capabilities to machines. Pattern is an entity, vaguely defined, that could be

given a name, e.g., fingerprint image, handwritten word, human face, speech signal, DNA sequence.

Character recognition is a sub-field of pattern recognition.

Various areas of Pattern Recognition such

- **Template matching.**
- **Statistical pattern recognition.**
- **Artificial Neural Networks.**

ANNs are often used in pattern recognition and machine learning. Despite having been used for decades, Feed-Forward Back-Propagation (FF-BP) systems are still the most commonly used ANN topology. FF-BP ANNs are applied in an extensive range of areas including computer network security, handwriting analysis, medicine, intrusion detection, computer vision, physics, retail, battlefield management, and finance. Their performance depends on several factors, including: the number of neuronal layers, the number of neurons at each layer, the activation functions used by the neurons, and the choice of initial connection weights. There are many algorithm for ANN learning algorithm such as Adaline, Hebbian , Perceptron Learning rule, Back propagation.

Yana M., Mohmad H. and Rozaida [1]. G. [2010], Training a Functional Link Neural Network Using an Artificial Bee Colony for Solving a Classification Problems. Faez H., Hazem N. [9], 2013, used Particle Swarm Optimization for training artificial neural network algorithm in Greek Letter Recognition. Sudarshan N., et al. [12], 2012, proposed method an improved artificial bee colony algorithm based back-propagation neural network training method for fast and improved convergence rate of the hybrid neural network learning method. Satish L., CH.V.Sarma[11],2013, proposed an efficient Hybrid Particle Swarm Optimization and Back propagation algorithm to enhance the performance of Artificial Neural Network for Handwriting Recognition by means of optimizing ANN weights.

In this paper, Artificial Bee Algorithm is used to learn the NN and the result are compared with BP method in order to specified which one is better in character recognition problems.

2. Artificial Neural Network

It is an information-processing system that has certain performance characteristics in common with biological neural networks. ANNs represent an important area of research, which opens a variety of new possibilities in different fields including classification or pattern recognition or predictions.

It is known that NN can approximate functions and mathematical operators arbitrarily as well as the number of neurons in the network tends to infinity. In this respect Feed Forward Artificial Neural Networks (FFANNs) can be considered as "universal approximations" which is capable of describing the input-output relationships of mechanical systems [3].

With classification in application domain of ANNs, there is a long list of researches and real applications include speed processing, image processing and computer vision, pattern classification and recognition, system control, robotics, forecasting and modeling, optimization and management of information and medical diagnosis [4]. Also there is a great role of NNs as a means for implementing expert systems, because of their ability to solve a specific problem, producing it as if it were a black-box solution where the mode of producing answers is not clearly understood [5]. Due to its adaptive and parallel processing ability, it has many applications in the engineering field [6].

NNs can be grouped into six areas of applications: prediction, pattern recognition, associative memories, classification, optimization and general mapping.= The back propagation learning algorithm can be divided into two phases: propagation and weight update.

Phase 1: Propagation

Each propagation involves the following steps:

1. Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.

2. Backward propagation of the propagation's output activations through the neural network using the training pattern target in order to generate the deltas of all output and hidden neurons.

Phase 2: Weight update

For each weight-synapse is multiply its output delta weight and input activation to get the gradient of the weight. Where delta is defined in equation (1)

$$\Delta w_i = \eta(t - o) x_i \quad \dots 1$$

Where η is learning rate, t is target, o output of net, x_i is input of net.

```

Initialize network weights (often small random values)
Do
  For Each training example ex
    Prediction = neural-net-output (network, ex) //
    forward pass
    Actual = teacher-output (ex)
    Compute error (prediction - actual) at the output
    units
    Compute  $\Delta w_h$  for all weights from hidden layer
    to output layer // backward pass
    Compute  $\Delta w_i$  for all weights from input layer to
    hidden layer //backward pass
    Update network weights
  Until all examples classified correctly or another
  stopping criterion satisfied
Return the network
    
```

behavior of honey bee colonies. The model consists of three essential components: employed and unemployed foraging bees, and food sources. The first two components, employed and unemployed foraging bees, search for rich food sources, which is the third component, close to their hive. The model also defines two leading modes of behavior which are necessary for self-organizing and collective intelligence: recruitment of foragers to rich food sources resulting in positive feedback and abandonment of poor sources by foragers causing negative feedback [2].

In ABC, a colony of artificial forager bees (agents) search for rich artificial food sources (good solutions for a given problem). To apply ABC, the considered optimization problem is first converted to the problem of finding the best parameter vector which minimizes an objective function. Then, the artificial bees randomly discover a population of initial solution vectors and then iteratively improve them by employing the strategies: moving towards better solutions by means of a neighbor search mechanism while abandoning poor solutions [11].The ABC algorithm shown in figure (1).

3. The Artificial Bee Colony (ABC)

The Artificial Bee Colony (ABC) algorithm is a swarm based meta-heuristic algorithm that was introduced by Karaboga in 2005 (Karaboga, 2005) for optimizing numerical problems. It was inspired by the intelligent foraging behavior of honey bees. The algorithm is specifically based on the model proposed by Tereshko and Loengarov (2005) for the foraging

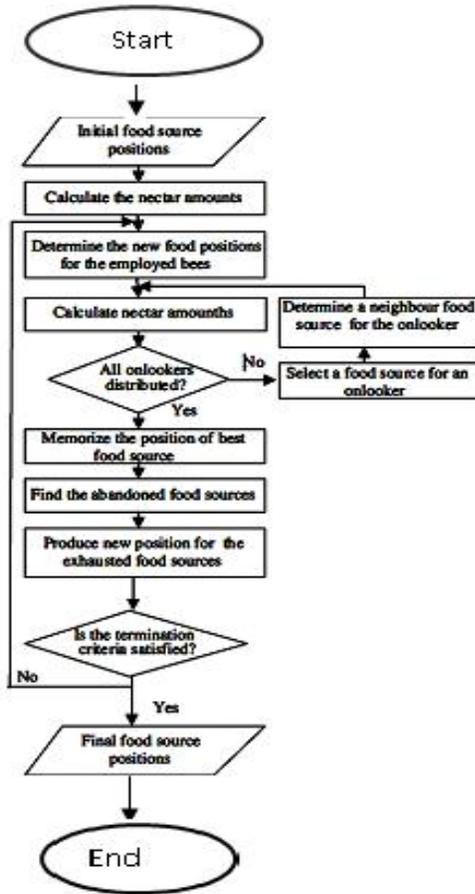


Figure 1: Flow chart of the ABC algorithm

3.1 The Artificial Bee Colony Algorithm

Detailed Pseudo code of the ABC Algorithm

- 1: Initialize the population of solutions $x_{i,j}$
- 2: Evaluate the population
- 3: cycle=1
- 4: repeat
- 5: Produce new solutions (food source positions) $v_{i,j}$ in the neighborhood of $x_{i,j}$ for the Employed bees using the formula $v_{i,j} = x_{i,j} + \Phi_{ij}(x_{i,j} - x_{k,j})$ (k is a solution in the Neighborhood of i , Φ is a random number in the range $[-1,1]$) and evaluate them
- 6: Apply the greedy selection process between x_i and v_i
- 7: Calculate the probability values P_i for the solutions x_i by means of their fitness values using the equation (2)

$$P_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i} \quad 2$$

In order to calculate the fitness values of solutions we employed the following equation (eq. 3):

$$fit_i = \begin{cases} \frac{1}{1+f_i} & \text{if } f_i \geq 0 \\ 1+abs(f_i) & \text{if } f_i < 0 \end{cases} \quad 3$$

Normalize P_i values into $[0,1]$

8: Produce the new solutions (new positions) v_i for the onlookers from the solutions x_i , selected depending on P_i , and evaluate them

9: Apply the greedy selection process for the onlookers between x_i and v_i

10: Determine the abandoned solution (source), if exists, and replace it with a new randomly produced solution x_i for the scout using the equation (4)

$$x_{ij} = \min_j + \text{rand}(0,1) * (\max_j - \min_j) \quad (4)$$

11: Memorize the best food source position (solution) achieved so far

12: cycle=cycle+1

13: until cycle= Maximum Cycle Number (MCN) [7,11 ,8].

3.2 Fitness Criterion

One of these stopping criteria is the fitness value. Since the BA algorithm is chosen to be a supervised learning algorithm, then there are observed values of (t_i) and desired output values of (f_i). These two values have to be compared, if they are closed to each other then the fitness is good, else the algorithm must continue its calculations until this condition is satisfied or the specified number of iterations is finished.

The corrections to the weights are selected to minimize the residual error between t_i and f_i output. The Mean Squared Error (MSE) is one solution for the comparison process used equation (5):

$$MSE = \frac{1}{n} \sum_{i=1}^n (t_i - f_i)^2 \quad 5$$

Where n is the number of the compared categories.

4. Training the ANN by using ABC

The ANN is statistical model of cognition that input vectors of independent variables and outputs estimates of vectors of dependent variable. The network is structure as set of weights, usually arranged in layers, and the optimization problem is to find values for the weight that make the mapping minimal error.

For the purpose of NN learning the mean square error referred to as the object function to be optimized by the optimization method. Each patch represent a candidate solution to the optimization problem, a patch represent the weight vector of NN including all biases.

Update position mean update the weight of network in order to reduce error. The new patch is set of new weight and biases used to obtain the new error.

The training process continues until satisfactory error is reached or computational limits are exceeded. When the training ends, the weights are used to calculate the classification error for the training patterns. The same set of weights is used then to test the network using the test patterns.

One of the drawbacks of the algorithm is the number of tunable parameters used. However, it is possible to set the parameter values by conducting a small number of trials. Other optimization algorithms usually employ gradient information. However, the proposed algorithm makes little use of this type of information and thus can readily escape from local optima.

The ABC algorithm is vastly different than any of the traditional methods of training. ABC does not just train one network, but rather training networks. ABC builds a set number of ANN and initializes all network weights to random values and starts training each one. On each pass through a data set, ABC compares each network's fitness. The network with the highest fitness is considered the global best [10].

Artificial Bee algorithm has been used in several optimization problems, including the optimization of synaptic weights from an Artificial Neural Network to generate a robust ANN. [11].

In this paper, the ABC was implemented in classification applications. From the classification application, the character recognition problem is chosen.

A NN-Learning system has been proposed to solve the mentioned problems by ABC and BP algorithms.

The proposed procedure for training ANN for character recognition showed in figure (2). The procedure start by reading input data of character is save as matrix and feed into ABC, ABC initializes all network weights to random values and continue ABC algorithm steps using mean square error as objective function until maximum number of iteration reached then NN weights and basises is save in matrices to use in test phase.

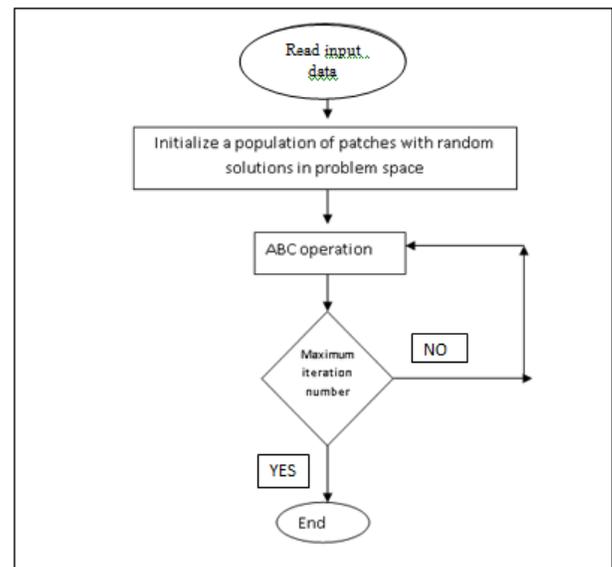


Figure 2: The proposed procedure of training ANN

The patterns are represented by Row *Column (RxC) matrix. The data of the matrix consists of 1's and 0's only. The objects of interest are the 1's and the background consists of 0's.

Where

- **Ns:** is the number of training set. **K** is number of character to recognized.

- **M** : number of neuron in hidden layer .

The proposed CR neural network consists of 3 layers, with $(R \times C, M, k)$ neurons which represent the number of neuron in input, hidden and output layers respectively. The proposed CR-NN is shown in figure (3).

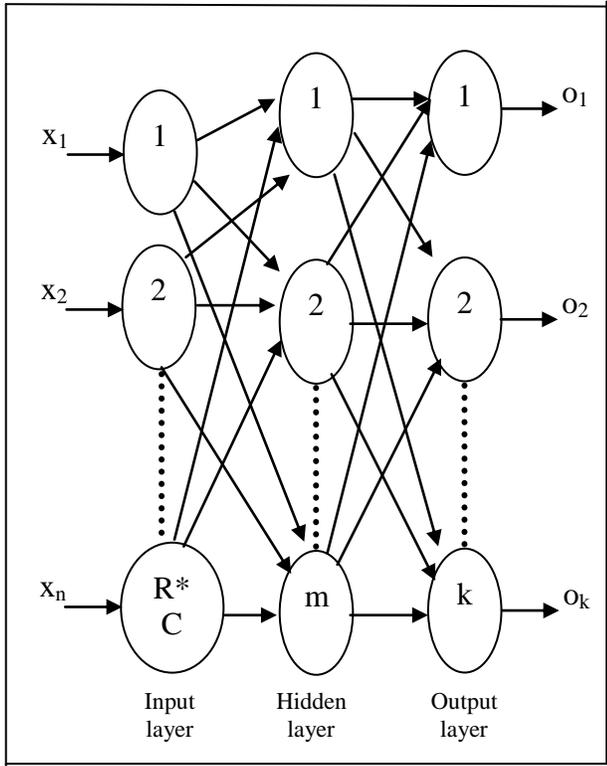


Figure 3: the proposed CR-NN $(R \times C, M, k)$.

First it has to choose random weights for the network interconnections ranging between 0 and 1.

Character Recognition has been defined as the conversion of text characters into machine readable codes .The aim was to recognize character pattern is represented by 5×5 matrix $(R=5, C=5)$, this means that (25) neurons are in the input layer. The training set $N_s=32=8 \times 4$, each character was represented by four matrices, one of them gives the correct shape, in other words, the noise in the input data is about 75% of the whole input data. Table (1) shows the eight characters binary representation.

8 neurons can be suggested to be used in the hidden layer.

The results of ABC implementation in $(25-8-8)$ CR-NN for 5 experiments are shown in table (3). We need to use the following symbols in the tables of ABC implementation:

- **Exp**: Experiment number.
- **NoI**: Number of Iterations.
- **MSE**: least fitness (Mean Square Error).
- **AL**: Accuracy Level of the learning

Table 1: show the eight characters (5×5) matrix representation

	+					-					×					÷				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0
2	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1
4	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0
	/					\										○				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	1	0
2	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1
3	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1
4	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1
5	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0

In Table (2) show the 8 character that used for CR

Table 2: the output binary coding of CR.

i	Pattern	Shape
1	Positive sign	+
2	Minus sign (Horizontal line)	-
3	Multiplication sign	×
4	Division sign	÷
5	Italic line from right to left	/
6	Italic line from left to right	\
7	Vertical line	
8	Circle	○

Experimental result of ABC algorithm in 5 times and corresponding mean square error and accuracy level is shown in table 3 .

Table 3: ABC implementation in (25-8-8) CR-NN.

Exp	NoI	MSE	AL %
1	193	0.10055	50.4%
2	354	0.07104	87.0%
3	719	0.05551	92.9%
4	2452	0.04196	93.3%
5	3285	0.03978	93.75%

6. Comparison Results between ABC and BP

The multi-layer perceptrons established for relevant purposes are trained with ABC algorithm and BP algorithm, respectively, through the selected training set. In order to establish a fair start-up state for BP-based perceptrons, the training processes of BP-based perceptrons always start with best solution in the initial population used for the processes of the counterparts, ABC-based perceptrons. For BP-based perceptrons, such “evolution time” directly equals the times of its updated iterations, while for BA-based perceptrons, it is equivalent to a production of the population size and the evolving generations. In order to compare performances between ABC and BP-based perceptrons, the training histories are recorded with the same fitness evolution times for both perceptrons to be compared with.

In this study, some examples of letter recognition are chosen, in order to make a comparison study for CR-NN between BA and BP algorithms.

In the CR-problems we chose $N_s=32$, for our comparative study between ABC and BP. The implementation is done for 5 experiments for each problem. Table (4) shows the comparison results. It’s obvious that BP is much better than ABC because of the thousands of iterations and bad accuracy level as shown in table (4).

Table 4: Comparison results between ABC and BP for CR-NN

Exp	Alg.	NoI	MSE	AL%
1	BA	194	0.101687	30.5%
	BP	142	0.101776	53.75%
2	BA	231	0.097150	72.3%
	BP	159	0.090614	60.2%
3	BA	327	0.076412	75.4%
	BP	172	0.070087	66.2%
4	BA	1070	0.050474	90.6%
	BP	327	0.031445	95.1%
5	BA	3999	0.038105	93.75%
	BP	625	0.000023	100.0%

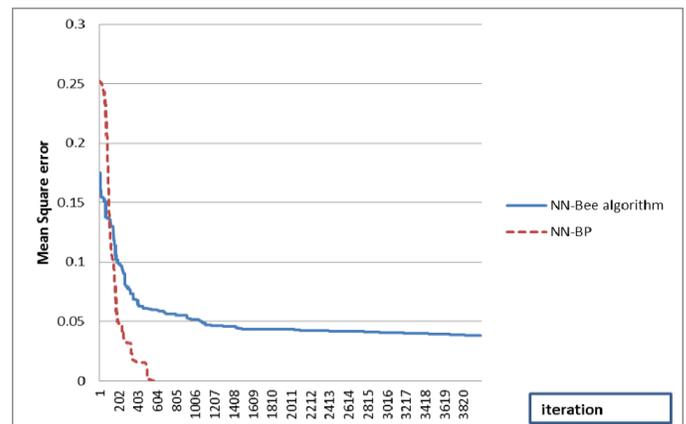


Figure 4: Show MSE value in each iteration for BP,ABC

Conclusions

From figure (4) concluded that BP algorithm reach to minimum error value faster than ABC algorithm, but BP algorithm have drawback is easily falling into local minima and fails to find the global optimum when used to train an NN, Because the error surface contains only a single global minimum, BP algorithm will converge to a weight vector with minimum error, regardless of whether the training examples are linearly separable while ABC algorithm, in spite of its population stochastic search method, it is less in falling in local minima .

From the result, The BP-based perceptron perform better than the ABC-based perceptron in application of pattern recognition. In other application such predication ABC gives performance better than BP [13]. In other word we can increase maximum iteration to reach acceptable error in exchange for sure that is optimal solution non sub-optimal solution and makes trade off.

References

- [1] Yana M., Mohmad H. and Rozaida G., " *Training a Functional Link Neural Network Using an Artificial Bee Colony for Solving a Classification Problems* "2010.
- [2] Bonabeau, E., Dorigo, M and Theraulaz, G., "Swarm Intelligence: From Natural to Artificial System", Oxford University Press, New York, USA. 1999.
- [3] Eskandarian A., and et al, "Vehicle Crash Modeling Using Recurrent Neural Networks", *Math. & Computer Modeling* vol. 28, No. 9, pp. 31 - 42, 1998.
- [4] Haykin S., " *Neural Network-a Comprehensive Foundation* ", prentice Hall, 1999.
- [5] Zurada J. M., " *Introduction to Artificial Neural Network* ", Jaico Publishing House, 1996.
- [6] Ahmed Faris, " *Image Compression Using Backpropagation Neural Network* ", M.Sc. thesis, University of Technology, Baghdad, 1995.
- [7] http://en.wikipedia.org/wiki/Artificial_bee_colony_algorithm.
- [8] <http://mf.erciyes.edu.tr/abc/>
- [9] Faez H., Hazem N., " *Pattern Recognition Based On Intelligent System* " Issue No. 31, *Journal of Al Rafidain University College* , ISSN (1681 – 6870) ,2013.
- [10] http://www.scholarpedia.org/article/Artificial_bee_colony_algorithm#Chong06
- [11] Pravin.Y. K and Shoba.K., " *Use Of Artificial Bee Colony (ABC) Algorithm in Artificial Neural Network Synthesis* ", *IJAEST*, Vol. 11, Issue No. 1,PP. 162 – 171,2010.
- [12] Sudarshan N., Partha P. and Achintya D., " training a feed forward neural network with artificial bee colony based back propagation ", *International Journal of Computer Science & Information Technology (IJCSIT)* Vol. 4, No 4,pp.33-46, August 2012.
- [13] http://www.scholarpedia.org/article/Artificial_bee_colony_algorithm.