ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

An Analysis of Mutation and Crossover in Evolutionary Neural Networks

Sunil Manohar Reddy K, Dr. G. Ravindra Babu, Dr. S. Krishna Mohan Rao

Assistant Professor, Department of Computer Science Engineering, Matrusri Engineering College, Hyderabad, Telangana, India

Professor, Department of Computer Science Engineering, Trinity College of Engineering & Technology, Karimnagar, Telangana, India

Professor, Department of Computer Science Engineering, Siddhartha Institute of Engineering & Technology, Hyderabad, Telangana, India

ABSTRACT: Evolving gradient-learning artificial neural network (ANNs) using an Evolutionary Algorithm (EA) is a popular approach to address the local optima and design problems of ANN. The typical approach is to combine the strength of Backpropagation (BP) in weight learning and EA's capability of searching the architecture space. However, the Backpropagation's "Gradient descent" approach requires a highly computer-intensive operation that relatively restricts the search converge of EA by completing it to use a small population size. To address this problem, we utilized mutation and crossover based genetic neural network to replace Backpropagation by using the mutation strategy of local adaption of Evolutionary strategy to affect weight learning.

The mutation and crossover enables the network to dynamically evolve its structure and adapt its weights at the same time. EP-based encoding scheme allows for a flexible and less restricted formulation of the fitness function and makes fitness computation fast and efficient. This makes it feasible to use larger population sizes and allows the mutation and crossover based genetic neural network to have relatively wide search coverage of the architecture space. Mutation and cross over genetic neural network implements a stopping criterion where over fitness occurrences are monitored through "Sliding Windows" to avoid premature learning and overlearning. Statistical analysis of its performance to some well known classification problems demonstrate its good generalization capability. It also reveals that locally adapting or scheduling the strategy parameters embedded in each individual network may provide a proper balance between the local and global searching.

KEYWORDS: Back propagation, Non-invasive, Artificial neural network, Mutation based genetic neural network.

I. INTRODUCTION

The use of evolutionary algorithms to aid in artificial neural networks learning has been a popular approach to address the shortcomings of back propagation. The typical approach uses a population of gradient-learning Artificial Neural Network (ANN) undergoing weight adaption through BP training and structure evolving through EA. At one extreme end are approaches that rely solely on EA for both of ANN's structure evolution and weight adaptation. Based on these observations and in relation to the proposed algorithm, these current approaches can be classified into two major types: "noninvasive" which refers to the former approaches where EA selection is used but fitness evolution requires BP or other gradient training; "invasive" which refers to the later approaches where the system uses EA for ANN's weight and structure evolution.

The proposed algorithm falls under the "invasive" approach. Since BP has been widely studied and many algorithms have been developed to improve its performance. The "noninvasive" evolution is the most popular. In this approach, the explicit separation of operations between weight adaptation by backpropagation and structure evolution by Evolutionary Algorithm often requires the development of a dual representation scheme. Because of this duality requirement, it is natural for this approach to adopt a GA-type evolution. The "noninvasive" approach does not change aggressively the typical learning mechanisms of the individual network. The Evolutionary Algorithm is only used to serve as a background process during evolution. Its successful performance still heavily relies on the proper initialization of backpropagation parameters and the proper choice of BP implementation. Individual network still undergoes gradient error minimization, which is prone to the "Local optima" problem.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

On the other hand, the "invasive" approach relies solely on an Evolutionary Algorithm for the ANN evolution. Since weight adaptation and structure evolution are carried out directly using Evolutionary Algorithm perturbation function, efficient implementation of this approach avoids the mapping problem by representing individuals at the species level. By using direct representation and avoiding back propagation fitness evaluation, important back propagation operations are fast and make feasible the use of a bigger population size for more robust search coverage. One area that needs attention for this approach to be successful is the development of appropriate encoding scheme that supports strong causality and evolvability, allows fast and efficient fitness evaluation, and provides facilities for simultaneous structure evolution and weight adaptation.

Another important area that needs to be addressed is the development of an appropriate stopping criterion to avoid premature learning and overlearning. One major issue of the "noninvasive" approach is the reliance to dual representations in representing BP-GA-aware ANNs. Finding a proper interpretation function to map ANN's genotype to phenotype is a big challenge due to the possible occurrence of deceptive mapping. Some of the major consequences of these deceptions include permutation problem or many-to-one mapping problem, one-to-many mapping problem, evolvability and causality problems.

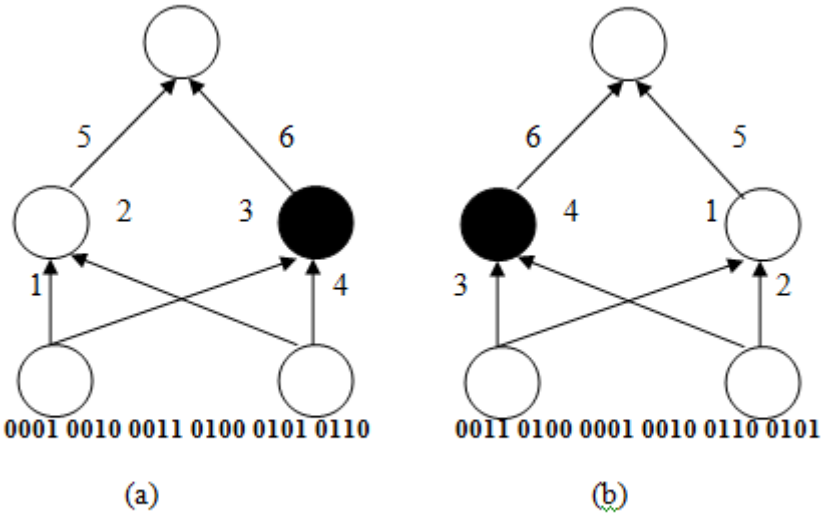

Figure 1: Deceptive mapping

Figure 1 illustrates an example of a competing convention problem or many-to-one mapping problem. Genotypes are represented by a string of binary digits where weights are encoded in chunks of 4 bits changes in the hidden layer due to crossover can produce two different genotypes with topologically similar ANN structure. Consequently, fitness evaluation in both structures produces the same output. This makes the evolution process inefficient and reduces population diversity which may lead to premature convergence.

In general, issues n causality, evolvability, deceptiveness, epitasis variance, ruggedness are consequences of this genotype-phenotype mapping. The intuitive notion of this issue is based on the principle that evolution and adaptations are possible if the improvement can be done in cumulative way or stepwise fashion. It is important that the mapping will not result into redundant phenotypes and inefficient fitness function evaluation. Also, the mapping must make sure that small changes in the genotype must have corresponding small changes to its phenotype, or else, evaluation becomes inefficient and hard to control evolvability and causality principles.

In the "noninvasive" approach, the high sensitivity of BP to the initial setup of its parameters causes a noisy evaluation of its fitness function which makes the entire process unreliable. Another problem is that BP is an iterative approach with no well-defined criterion to stop training. Since individuals undergo BP training for every generation to determine their fitness, the evolution compounds the inefficiency of gradient training.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

Due to this inefficiency, typical implementation prefers small population size to have faster convergence at the expense of having poor coverage of the search space. Hence, it will often miss the global solution and converges to local solution. On the other hand, typical fitness evaluation in the "Invasive" approach utilizes a one-way feed forward computation which is many times faster than backpropagation. Consequently, typical implementation prefers large population size because it has no significant impact in the overall CPU-time execution compared with that of the backpropagation implementation. Since "invasive" approach scales better than its counterpart, it has better potential to find the global solution.

II. EVOLUTIONARY STRATEGY OF TRAINING ARTIFICIAL NEURAL NETWORK

Analysis of weight and structure

Earliest implementations can be classified into three major approaches, namely: genetic algorithm (GA) Evolutionary Strategies (ESs) and evolutionary programming (EP). One popular EA approach to solve optimization problems is the GA. Holland's Scheme Theorem formalized GA and provided the theoretical underpinning of the mechanism behind it. One attractive feature of GA evolution is its support for the generic implementation of its major operations such as crossover, mutation, selection, and replacement. This is achieved by using a dual representation where evolutionary operations are done at the genotype level while fitness evaluation is carried out at the phenotype level.

On the other hand, earliest implementation of EP and ES relies on Gaussian mutation alone as the main source of perturbation. Instead of binary representation, both approaches were optimized to handle problems involving real-valued parameters. Notably, EP emphasizes the preservation of the behavioural links between parent and child. To carry out this objective, EP uses direct representation by representing individuals at the species level.

Moreover, it avoids using crossover operation to avoid the possibility of destroying the behavioural links from the parent to the child. The recent trend among practitioners is to combine these three different approaches to address more complicated problems in AI and other fields. To carry out the EA operations in ANN, it is important to have a proper encoding scheme provides high flexibility in problem representation but may reduce EA's efficiency due to complicated operations.

On the other hand, too simple representation may suffer from slow or premature convergence. This requires a careful selection of an encoding scheme such that EA operations are not compromised but still provides enough flexibility to support dynamic weight adaptation and evolution of structures.

Fig. 2 shows a typical example of a MGNN structure [Fig. 2(b)] and its equivalent ANN structure [Fig. 2(a)]. Individuals are represented by vectors and matrices of real numbers. Connection weights from the input layer to the hidden layer are encoded in the weight matrix W1 while the weight matrix W2 contains the connection weights from the hidden-layer to the output layer. These vectors and matrices are subjected to random perturbations during mutation to improve the ANN fitness.

One nice feature of this representation is its implicit support to structure evolution and weight adaptation. To illustrate, any column W2 in with values all set to zero represents a deleted or unutilized node in the hidden layer. For example, column W3 in W2 [Fig. 2(b)] indicates that there is no connection between node W3 [Fig. 2(a)] in the hidden layer to the output layer of ANN. This simple representation of ANN enables mutation and crossover of W1 and W2 to dynamically change the structure and weights of the network.
Dynamic structure changes and local adaptation of weights through mutation are implemented using

$$\delta = \rho(\sigma)$$
$$m^1 = m + \rho(\delta)$$
$$w^1_{ij} = w_{ij} + \rho\,(m^1)$$

where

$\sigma \rightarrow$ SSP (step size parameter)
$\delta \rightarrow$ mutation strength intensity
$\rho \rightarrow$ perturbation function
$m \rightarrow$ adapted strategy parameter

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

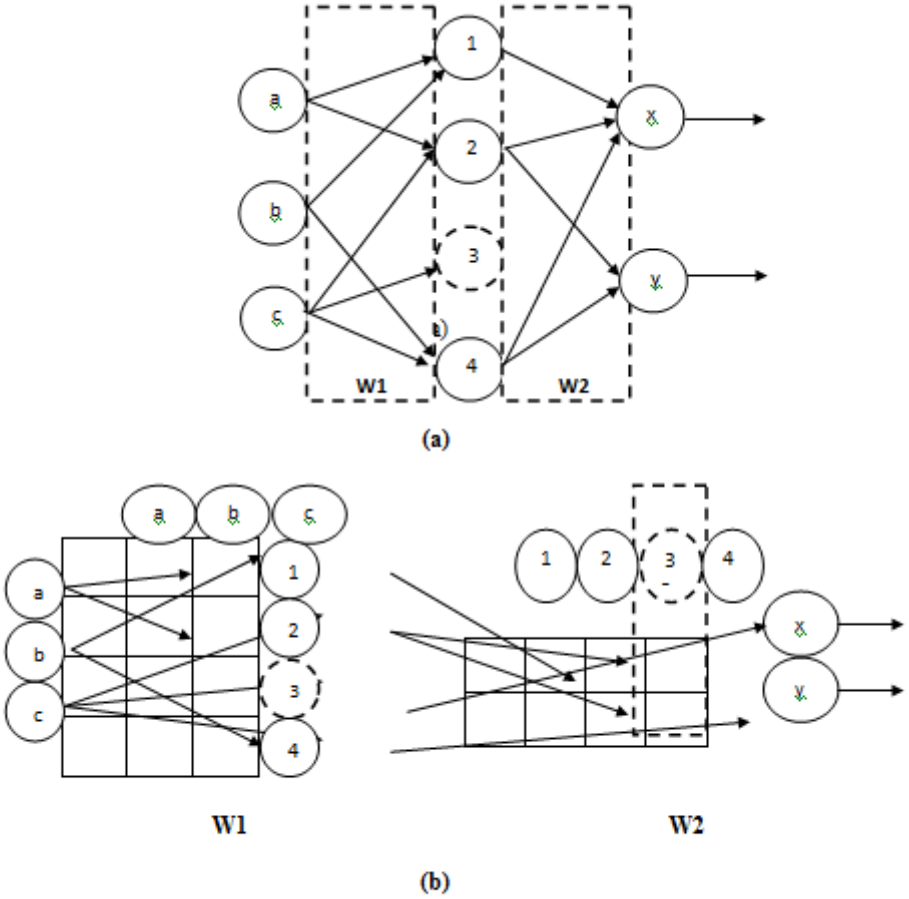

Figure 2: MGNN Encoding a) ANN b) MGNN

The maximum strength of mutation is controlled by the step size parameter (SSP) ϭ while the level (δ) of mutation strength is dynamically computed during evolution. Evolution is carried out by starting a population of networks with zero weights. Hence, the only way for any network to evolve is by mutation. Assuming that the distribution of ρ is symmetric, deletion of connections/nodes is implicitly done when the result of operation in is equal or closes to zero. Gaussian perturbations are concentrated in the neighbourhood of zero. MGNN uses this perturbation together with a small mutation probability to support strong causality and evolvability.

Data Set Information

UCI Repository

The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms.

Iris

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of Iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

Predicated attribute: class of iris plant.

This data differs from the data presented in Fishers article identified by Steve Chadwick, the 35th sample should be : 4.9,3.1,1.5,0.2,"Iris-setosa" where the errors are in the fourth feature. The 38th sample : 4.9,3.6,1.4,0.1, "Iris-setosa" where the errors are in the second and third features.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

Attribute Information

1. Sepal length in cm
2. Sepal width in cm
3. Petal lengthin cm
4. Petal width in cm
5. Class:

- Iris Setosa
- Iris Versicolour
- Iris Virginica
-

Dataset for Iris Setosa

5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa

Data Set Information for cancer

This data was used by Hong and Young to illustrate the power of the optimal discriminate plane even in ill-posed settings. Applying the KNN method in the resulting plane gave 77% accuracy. The data described 3 types of pathological lung cancers. The Authors give no information on the individual variables nor on where the data was originally used.

Dataset for lung cancer

1,0,3,0,?,0,2,2,2,1,1,1,1,3,2,2,1,2,2,
0,2,2,2,2,1,2,2,2,3,2,1,1,1,3,3,2,2,1,
2,2,2,1,2,2,2,2,2,2,2,2,2,2,1,1,1,2,2,
1,0,3,3,1,0,3,1,3,1,1,1,1,1,3,3,1,2,2,
0,0,2,2,2,1,2,1,3,2,3,1,1,1,3,3,2,2,2,
1,2,2,2,1,2,2,1,2,2,2,2,2,2,2,2,1,2,2,
1,0,3,3,2,0,3,3,3,1,1,1,0,3,3,3,1,2,1,
0,0,2,2,2,1,2,2,3,2,3,1,3,3,3,1,2,2,1,
2,2,2,1,2,2,1,2,2,2,2,2,2,2,2,2,1,2,
1,0,2,3,2,1,3,3,3,1,2,1,0,3,3,1,1,2,2,
0,0,2,2,2,2,1,3,2,3,3,1,3,3,3,1,1,1,1,
2,2,2,2,1,2,2,2,1,2,2,2,2,2,2,2,2,2,2,
1,0,3,2,1,1,3,3,3,2,2,2,1,1,2,2,2,2,2,
0,0,2,2,2,1,1,2,3,2,2,1,1,1,3,2,1,2,2,

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

1,2,2,2,1,2,2,2,2,2,2,2,2,2,2,1,2,2,
1,0,3,3,2,0,3,3,3,1,2,2,0,3,3,3,2,2,1,
0,0,1,2,2,2,1,3,3,1,2,2,3,3,3,2,1,2,2,
1,2,2,2,1,2,2,2,2,2,2,2,2,2,2,2,1,2,
1,0,3,2,1,0,3,3,3,1,2,1,2,3,3,3,3,2,2,
0,0,2,2,2,2,1,3,2,2,2,2,3,3,3,2,1,1,2,
2,1,2,1,2,2,2,2,1,2,2,2,2,1,2,2,2,1,2,
1,0,2,2,1,0,3,1,3,3,3,3,2,1,3,3,1,2,2,
0,0,1,1,2,1,2,1,3,2,1,1,3,3,3,2,2,1,2,
1,2,2,1,2,2,2,1,2,2,2,1,2,2,2,2,1,2,2,
1,0,3,1,1,0,3,1,3,1,1,1,3,2,3,3,1,2,2,
0,0,2,2,2,1,2,1,2,1,1,1,3,3,3,3,2,2,1,
2,2,2,1,2,2,1,2,2,2,2,2,2,2,2,1,2,2,
2,0,2,3,2,0,2,2,2,1,2,2,2,2,2,2,1,2,2,

Diabetes

Diabetes files consist of four fields per record.

4.4,2.9,1.4,0.2,
4.9,3.1,1.5,0.1,
5.4,3.7,1.5,0.2,
4.8,3.4,1.6,0.2,
4.8,3.0,1.1,0.1,
4.3,3.0,1.1,0.1,
5.8,4.0,1.2,0.2,
5.7,4.4,1.5,0.4,
5.4,3.9,1.3,0.4,
5.1,3.5,1.4,0.3,
5.7,3.8,1.7,0.3,

Thyroid disease

Dataset is identify the type of thyroid disease in the patient.

2.5,125,1.14,109	SVI,negative
2,102,0.98,109,0.91,120	SVI,negative
6,1.9,175,0.72,1.2,61,0.87,70	SVI,negative
3,183,1.3,141,2.2,0.6,80,0.7,15	SVI,negative
1.6,83,0.89,93	SVI,negative
2.2,115,0.95,121,	SVI,negative
3,1.8,109,0.91,119,	SVHC,negative
1,2.6,121,0.94,130	SVHC,negative
8,2.5,147,1.13,129	SVHC,negative
2.2,t,83,1.03,81	SVHC,negative
2,0.8,101,0.99,130	SVHC,negative
2.3,70,0.99,71	SVHC,negative
2.1,91,0.99,71	SVHC,negative
83,0.85,97	SVHC,negative
3,2,95,0.99,96	STMW, negative
7,125,1.43,87,	STMW, negative
147,0.95,154	STMW, negative
45,3.2,130,1.83,71	STMW, negative
9,3.3,156,1.67,93,	STMW, negative

Glass Identification

The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence, if it is correctly identified.

1.52101,13.64,4.49,1.10,71.78,0.06,8.75,0.00,0.00,1 building_windows _float_processed
1.51761,13.89,3.60,1.36,72.73,0.48,7.83,0.00,0.00,1 building_windows _float_processed
1.51618,13.53,3.55,1.54,72.99,0.39,7.78,0.00,0.00,1 building_windows _float_processed

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

1.51766,13.21,3.69,1.29,72.61,0.57,8.22,0.00,0.00,1 building_windows _float_processed
1.51742,13.27,3.62,1.24,73.08,0.55,8.07,0.00,0.00,1 building_windows _float_processed
6,1.51596,12.79,3.61,1.62,72.97,0.64,8.07,0.00,0.00,1building_windows float_processed
7,1.51743,13.30,3.60,1.14,73.09,0.58,8.17,0.00,0.00, 1building_windows non_float _processed
8,1.51743,13.30,3.60,1.14,73.09,0.58,8.17,0.00,0.00,1 building_windows non_float _processed
9,1.51756,13.15,3.61,1.05,73.24,0.57,8.24,0.00,0.00,1building_windowsnon_float_processed
10,1.51755,13.00,3.60,1.36,72.9,0.57,8.40,0.00,0.00,11,1building_windows_non_float_processed
11,1.51571,12.72,3.46,1.56,73.20,0.67,8.09,0.00,0.24,1building_windows _non_float_processed
12,1.51763,12.80,3.66,1.27,73.01,0.60,8.56,0.00,0.00,1building_windows _non_float_processed
13,1.51589,12.88,3.43,1.40,73.28,0.69,8.05,0.00,0.24,1vehicle_windows _non_float_processed
14,1.51748,12.86,3.56,1.27,73.21,0.54,8.38,0.00,0.17,1vehicle_windows_ float_processed
15,1.51763,12.61,3.59,1.31,73.29,0.58,8.50,0.00,0.00,1vehicle_windows _ float_processed
16,1.51761,12.81,3.54,1.23,73.24,0.58,8.39,0.00,0.00,1vehicle_windows_ float_processed
17,1.51784,12.68,3.67,1.16,73.11,0.61,8.70,0.00,0.00,1 vehicle _windows _ float_processed
18,1.52196,14.36,3.85,0.89,71.36,0.15,9.15,0.00,0.00,1 vehicle_windows _ float_processed

Water treatment plant

This dataset comes from the daily measures of sensors in a urban waste water treatment plant. The objective is to classify the operational state of the plant in order to predict faults through the state variables of the plant at each of the stages of the treatment process.

1.50,7.8,407,166,66.3,4.5,2110,7.9,228,70.2, ZN-E(input Zinc to plant)
5.5,2120,7.9,280,94,72.3,0.3,2010,7.3,84,21, ZN-E(input Zinc to plant)
81.0,0.02,2000,?,58.8,95.5,70.0,79.4,87.3, ZN-E(input Zinc to plant)
99.6,39024,3.00,7.7,443,214,69.2,6.5,2660, ZN-E(input Zinc to plant)
1666,7.7,220,72.7,4.5,1594,7.7,272,92,78, PH-E(input pH to plant)
1742,7.6,128,21,81,0.05,1888,58.2,95.6,52.9, PH-E(input pH to plant)
75.8,88.7,98.5,35023,3.50,7.9,205,588,192,65.6, PH-E(input pH to plant)
4.5,2430,7.8,236,268,73.1,8.5,2280,7.8,158,376, PH-E(input pH to plant)
236,57.6,4.5,2020,7.8, DBO-E(input Biological demand of oxygen to plant)
372,88,68.2,0.2,2250, DBO-E(input Biological demand of oxygen to plant)
7.6,19,108,22,65.9, DBO-E(input Biological demand of oxygen to plant)

III. PROPOSED ALGORITHMS

Exemplar based Inpainting technique is used for inpainting of text regions, which takes structure synthesis and texture

Algorithm : Training Algorithm

Data : parents={net_1, ……$net_{popsize}$};
Nextparents;window;fittnessNet; betsNet;
Result : bestNet
Begin
Parents.initilize();
Parents.crossover();
For generation ← 0 to maxGeneration do
¥ net ЄParents,net.coputeFitness();
If fmod(generation,windowsize)==0 then
If stop(fitestNet,bestNet) then break;
nextParents ← select (parents);
swap(Parents,next parents);
parents.mutate();
return bestNet;
end

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

Algoritham : Crossover with Stichastic Policy

Data : parents={net_1, ……$net_{popsize}$};
Result : crossover_parents
Begin
For each network net Є Parents,net !=fittnessNet
Do /* compute crossover to be adapted*/
Net.crossover+=gaussProb(0,net.crssover);
For each weight matrix W Є net do /* crossover weight matrices*/
For i←to rowSize of net.W do
For j← to columnsize of net. W do
If uniformRandom(0,1) ≤
Net.crossoverProb then
Net[i][j]+=Gaussprob(0,net.Crossover);
For each threshold vector Є net do /* crossover Vectors*/
If uniformRandom(0,1) ≤
Net.crossoverProb then
Net.T[i][j]+=
Gaussprob(0,net.Crossover);
End

IV. IMPLEMENTATION

All the experiments are conducted on a HCL desktop, Intel ®Pentium ® D CPU 3.20 GHz with 1 GB of RAM running Fedora 8 Linux operating system. All the algorithms are implemented in GNU C++.
Mutation and Crossover are the methods used to classify the data sets. In crossover three methods are used one point crossover, twopoint crossover and uniform crossover. The datasets for Iris, Lung Cancer, Diabetes, Thyroid diseases, Glass identification, Water Treatment Plant are taken from UCI repository. Rank based selection and roulette wheel selection are used to classify the datasets. According to the experimental results obtained Rank based selection gave better results comparing with Roulette wheel selection. Classification of data is represented in form of tables, graphs.

V. EXPERIMENTS AND RESULTS

Preliminary experiments showed that mutation, crossover and selection probability of 0.01 significantly outperformed mutation probability of 0.05. These experiments indicated the superior performance of using Gaussian perturbation over uniform perturbation. Consequently, all simulations used Gaussian mutation with the mutation probability fixed at 0.01 to evaluate the performance several experiments were conducted using some well-known classification problems from the database found in the UCI repository: iris classification data, Diabetes recognition data, Cancer classification data, Thyroid diseases data , Glass identification data, water treatment plant data.
The effects of these factors to the performance of MGNN were measured using the following dependent variables:

- Percentage of wrong classification(Class Errors);
- Number of connection weights(connections);
- Number of generations (generations).

The implementations use a combination of training and validation data during the training phase, Mutation based genetic neural network (MGNN) variants strictly use separate datasets for each phase and do not allow intersection. The data were divided into 50% training, 25% validation, and 25% testing using SRS (Simple Random Sampling) expect for the cancer problem. This allows statistically valid comparisons of MGNN with other algorithms that use the same datasets. Unfortunately, these standard sets are limited in scope and do not include the diabetes and iris problems. In the object-oriented paradigm, roulette wheel method serves as the base class with rank-based selection inheriting all of the former's features except the selection policy.
The sizes for the input and output units are problem-specific while the maximum number of hidden units is a user-defined parameter. The classification method uses 1-of-m encoding for classes using output values of either 1 or 0. One generation in Mutation and crossover is equivalent to a single presentation of the entire training data to each member of the population. For example, using a population size of 100 implies that one generation is roughly equivalent to 100 epochs without error. All variants use a strong restriction in the classification policy by considering correct

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

classification only if the maximum absolute error is less than or equal to 0.3 (i.e., Max ($|T_i-O_i|$)). Majority of these features can be tweaked to improve MGNN performance.

Training Performance

This trend is fairly consistent to all the three problems. Since the performance of the three variants are optimal in the cancer problem. The discrepancy in performance among SSPs is not as apparent as the other two problems shows the training performances based on the percentage of correct classification. In just less than 150 generations. In particular, it correctly classified more than 90% of the cancer data in less than 50 generations. Both rank-based selection and roulette wheel selection took less than 300 generations to classify 90% of data in all problems.

Testing Performance

One of the most important criteria to determine the effectiveness of ANN learning is its generalization capability when confronted with a completely new set of data.

Table 1: Testing Performances

Strategy		*Test error*	*Connections*	*Generations*
Cancer:	Rank	3.1	66	113
	Roul	4.	67.1	267
Iris:	Rank	6.1	68.2	403
	Roul	7.1	70.4	617
Diabetes:	Rank	8.1	72.3	804
	Roul	9.1	73.2	941
Thyroid disease:	Rank	3.1	66	113
	Roul	4.	67.1	267
Glass Identification:	Rank	6.1	68.2	403
	Roul	7.1	70.4	617
Water Treatment Plant:	Rank	8.1	72.3	804
	Roul	9.1	73.2	941

To aid in the analysis of the generalization performance, DMRT tables are presented to highlight significant differences in the performance of the different variants. Mean values of factors in different subsets (columns) indicate that they have significantly different performances at 0.05 level of significance. Table 1 shows the generalization performances of the different variants. These were computed based on their average performances over all SSPs using a total of 150 trails. The table indicates that the generalization performances of the three approaches are not significantly different in the cancer problem. For the iris and diabetes problems, rank-based performed better than roulette-wheel.

In addition, the table below shows that rank-based selection significantly used less number of connections compared to the other two variants. The table also indicates that the scheduled mutation strategy improved the generalization performance of roulette wheel over rank based selection at the cost of significantly using more connections. However, this added complexity did not negatively affect the generalization performance of rank based. Except for the diabetes problem, there is no significant difference in the average number of generations among all variants.

Roulette wheel significantly used the least number of generations. It significantly exhibited the worst generalization performance (only 8.46%). It also performed worst during the training phase. The table indicates that there is no significant difference in the generalization performances of the three step size parameters. All variants require an average of less than 250 generations to achieve about 3% errors in the testing data of the cancer problem. For the iris problem, rank based had significantly better classification (4.68%) than MGNN-rank (7.46%) and roulette (7.35%).

All variants required an average of less than 450 generations and an average of less than 55 connections. The diabetes problem required all variants to have relatively bigger number of connections and more generations compared to the other two problems. With the exception of and roulette wheel, rank based selection and relatively small classification error in spite of the added complexity and generations.

During training, the rank-based selection policy was a better option than the fitness-based selection policy. Also, the scheduled mutation policy was a better option than the ordinary stochastic mutation policy. Rank based selection having both of these features consistently outperformed the other approach in the three classification problems. The trend in overfitness and correlation of Backpropagation variants with good solutions seems to indicate that a relatively high over fitness value has a corresponding high correlation value. In addition, it suggests that a small over fitness in the validation data can be useful in the generalization.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

Tolerating over fitness, however, may have a bad consequence if applied to other noisy problems. Variants with good training performances have also good generalization performances. One way to achieve this is to use large SSP. Large SSP speeds up convergence and minimizes network complexity. On the other hand, using a small SSP provides similar generalization performance, able it much slower and requires significantly more connections. This slow convergence, however, may be useful in the later part of evaluation to refine the search. This suggests that adapting or annealing step size parameter may help to improve the overall performance. Developing an annealed or scheduled Step Size Parameter (SSP) will be further investigated in the future.

Small step size parameter (SSP) causes good correlation in training, validation, and testing. However, it produces networks with poor training and generalization performances. Large SSP produces the opposite trend, i.e., poor correlations but superior performance in training and generalization. Using small SSP allows the network to converge to common local minima in the training, validation, and testing data. Consequently, this causes high correlations. On the other hand, using large SSP enables the network to escape these local minima but needs smaller SSP at the later part of the evaluation to localize its search and improve the correlation.

VI. TABLES

TABLE 2: MUTATION FOR IRIS

Generations	Rankbased selection	Roulette wheel selection
251	0.55	0.0209
271	0.122	0.0496
275	0.188	0.0748
511	0.256	0.244733
572	0.32	0.264
694	0.456	0.498
818	0.522	0.5778
861	0.589	0.676
873	0.656	0.797

This table gives the result of Mutation for Iris classification. Rank based Selection and roulette wheel selection is used to classify Iris dataset. Roulette wheel selection gives better result than Rank based selection.

TABLE 3: MUTATION FOR CANCER

Generations	Rankbased selection	Roulette wheel selection
251	0.211	0.0209
271	0.277	0.0496
275	0.344	0.0748
511	0.477	0.244733
572	0.544	0.264
694	0.61	0.498
818	0.678	0.5778
861	0.744	0.676
873	0.811	0.797

This table gives the result of Mutation for Cancer classification. Rank based Selection and roulette wheel selection is used to classify Cancer dataset. Rank based selection gives better result than roulette wheel selection.

Table 4: Mutation for Diabetes

Generations	Rankbased selection	Roulette wheel selection
251	0.333	0.0209
271	0.122	0.0496
275	0.211	0.0748
511	0.3	0.244733
572	0.478	0.264
694	0.567	0.498
818	0.656	0.5778
861	0.744	0.676
873	0.833	0.797

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

This table gives the result of Mutation for diabetes classification. Rank based selection and roulette wheel selection is used to classify Diabetes dataset. Rank based selection gives better result than roulette wheel selection.

Table 5: Onepoint Crossover for Iris

Generations	Rankbased selection	Roulette wheel selection
121	0.556	0.02
173	0.211	0.74
297	0.288	0.153
317	0.366	0.21
434	0.444	0.37
538	0.522	0.40
598	0.6	0.0550
703	0.6778	0.682
763	0.756	0.758

This table gives the result of Onepoint crossover for Iris classification. Rank based selection and roulette wheel selection is used to classify Iris dataset. Roulette wheel selection gives better result than Rank based selection.

Table 6: One point Crossover for Cancer

Generations	Rankbased selection	Roulette wheel selection
121	0.033	0.02
173	0.211	0.74
297	0.3	0.153
317	0.477	0.21
434	0.567	0.37
538	0.656	0.40
598	0.744	0.0550
703	0.833	0.682
763	0.91	0.758

This table gives the result of Onepoint crossover for Cancer classification. Rank based selection and roulette wheel selection is used to classify Cancer dataset. Rank based selection gives better result than Roulette wheel selection.

Table 7: One point Crossover for Diabetes

Generations	Rankbased selection	Roulette wheel selection
121	0.077	0.02
173	0.211	0.74
297	0.278	0.153
317	0.411	0.21
434	0.478	0.37
538	0.544	0.40
598	0.611	0.0550
703	0.677	0.682
763	0.744	0.758

This table gives the result of Onepoint crossover for Diabetes classification. Rank based selection and roulette wheel selection is used to classify Diabetes dataset. Roulette wheel selection gives better result than Rank based selection.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

Table 8: Uniform Crossover for Iris

Generations	Rankbased selection	Roulette wheel selection
121	0.36	0.02
242	0.44	0.074
363	0.522	0.153
484	0.6	0.219
605	0.677	0.376
726	0.755	0.40
847	0.833	0.0550
968	0.911	0.68
1089	0.988	0.758

This table gives the result of Uniform crossover for Iris classification. Rank based selection and roulette wheel selection is used to classify Iris dataset. Rank based selection gives better result than Roulette wheel selection.

Table 9: Uniform Crossover for Cancer

Generations	Rankbased selection	Roulette wheel selection
121	0.322	0.02
242	0.377	0.074
363	0.433	0.153
484	.489	0.219
605	0.54	0.376
726	0.6	0.40
847	0.711	0.0550
968	0.767	0.68
1089	0.822	0.758

This table gives the result of Uniform crossover for Cancer classification. Rank based selection and roulette wheel selection is used to classify cancer dataset. Rank based selection gives better result than Roulette wheel selection.

Table 10: Uniform Crossover for Diabetes

Generations	Rankbased selection	Roulette wheel selection
121	0.266	0.02
242	0.489	0.074
363	0.54	0.153
484	0.6	0.219
605	0.656	0.376
726	0.711	0.40
847	0.767	0.0550
968	0.822	0.68
1089	0.878	0.758

This table gives the result of Uniform crossover for Diabetes classification. Rank based selection and roulette wheel selection is used to classify Diabetes dataset. Rank based selection gives better result than Roulette wheel selection.

Table 11: Mutation for Thyroid Classification

Generations	Rankbased selection	Roulette wheel selection
251	0.556	0.496
271	0.211	0.244
275	0.288	0.64
512	0.366	0.268
694	0.444	0.55
818	0.522	0.67
861	0.6	0.79
873	0.756	0.839

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

This table gives the result of Mutation for Thyroid classification. Rank based selection and roulette wheel selection is used to classify Thyroid dataset. Roulette wheel selection gives better result than Rank based selection.

Table 12: Mutation for Glass Identification

Generations	Rankbased selection	Roulette wheel selection
251	0.077	0.496
271	0.133	0.244
275	0.188	0.64
512	0.358	0.268
694	0.578	0.55
818	0.633	0.67
861	0.8	0.79
873	0.858	0.839

This table gives result of Mutation for Glass identification classification. Rank based selection, roulette wheel selection is used to classify Glass identification dataset. Rank based selection gives better result than Roulette wheel selection.

Table 13: Mutation for Water Treatment Plant

Generations	Rank based selection	Roulette wheel selection
251	0.44	0.496
271	0.211	0.244
275	0.344	0.64
512	0.411	0.268
694	0.611	0.55
818	0.811	0.67
861	0.8778	0.79
873	0.944	0.839

This table gives the result of Mutation for Water Treatment Plant classification. Rank based selection and roulette wheel selection is used to classify Water Treatment Plant dataset. Rank based selection gives better result than Roulette wheel selection.

Table 14: Two point Crossover for Thyroid disease

Generations	Rankbased selection	Roulette wheel selection
121	0.077	0.02
173	0.211	0.74
297	0.278	0.153
317	0.411	0.21
434	0.478	0.37
538	0.544	0.40
598	0.611	0.55
703	0.677	0.682

This table gives the result of Twopoint crossover for thyroid classification. Rank based selection and roulette wheel selection is used to classify Thyroid dataset. Roulette wheel selection gives better result than Rank based selection.

Table 15: Twopoint Crossover for Glass identification

Generations	Rankbased selection	Roulette wheel selection
121	0.36	0.02
173	0.44	0.74
297	0.522	0.153
317	0.6	0.21
434	0.677	0.37
538	0.755	0.40
598	0.833	0.55
703	0.911	0.682

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 2 , February 2016

This table gives the result of Twopoint crossover for Glass identification classification. Rank based selection and roulette wheel selection is used to classify Glass identification dataset. Rank based selection gives better result than Roulette wheel selection.

Table 16: Twopoint Crossover for Water Treatment Plant

Generations	Rankbased selection	Roulette wheel selection
121	0.44	0.02
173	0.522	0.74
297	0.6	0.153
317	0.677	0.21
434	0.755	0.37
538	0.833	0.40
598	0.911	0.55
703	0.989	0.682

This table gives the result of Twopoint crossover for Water Treatment Plant classification. Rank based selection and roulette wheel selection is used to classify Water Treatment Plant dataset. Rank based selection gives better result than Roulette wheel selection.

Table 17: Uniform Crossover for Thyroid disease

Generations	Rankbased selection	Roulette wheel selection
121	0.095	0.02
242	0.18	0.074
363	0.265	0.153
484	0.35	0.219
605	0.43	0.376
726	0.52	0.40
847	0.69	0.550

This table gives the result of Uniform crossover for Thyroid disease classification. Rank based selection and roulette wheel selection is used to classify Thyroid dataset. Rank based selection gives better result than Roulette wheel selection.

Table 18: Uniform Crossover for Glass identification

Generations	Rankbased selection	Roulette wheel selection
121	0.556	0.02
242	0.211	0.074
363	0.288	0.153
484	0.366	0.219
605	0.444	0.376
726	0.522	0.40
847	0.6	0.550

This table gives the result of Uniform crossover for Glass identification classification. Rank based selection and roulette wheel selection is used to classify Glass identification dataset. Rank based selection gives better result than Roulette wheel selection.

Table 19: Uniform Crossover for Water Treatment Plant

Generations	Rankbased selection	Roulette wheel selection
121	0.289	0.02
242	0.367	0.074
363	0.44	0.153
484	0.678	0.219
605	0.756	0.376
726	0.833	0.40
847	0.911	0.550

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 2 , February 2016

This table gives the result of Uniform crossover for Water Treatment Plant classification. Rank based selection and roulette wheel selection is used to classify Water Treatment Plant dataset. Rank based selection gives better result than Roulette wheel selection.

VII. CONCLUSION

The mutation and crossover enables the network to dynamically evolve its structure and adapt its weights at the same time. EP-based encoding scheme allows for a flexible and less restricted formulation of the fitness function and makes fitness computation fast and efficient. This makes it feasible to use larger population sizes and allows the mutation and crossover based genetic neural network to have relatively wide search coverage of the architecture space. Mutation and cross over genetic neural network implements a stopping criterion where over fitness occurrences are monitored through "Sliding Windows" to avoid premature learning and overlearning.

REFERENCES

[1] Werbos, P et al.: The Roots of Back propagation, From Ordered Derivatives to Neural Networks and Political Forecasting. John Wiley, New York, NY, 1993.
[2] Plaut et al.: Experiments on learning by back propagation (CMU-CS-86-126). Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA, Prentice Hall, India, 1998.
[3] Franzini MA et al.: Speech recognition with back propagation. Proceedings of the IEEE/Ninth Annual Conference of Engineering in Medicine and Biology Society, Boston.MA, 9, 1702-1703, 1987.
[4] Chen J R et al.: Step size variation methods for accelerating the back propagation algorithm. Proceedings of the International Joint Conference on Neural Networks, Washington, DC, 601-604, 1990.
[5] Holt et al.: Convergence of back propagation in neural networks using a log-likelihood cost function. Electron Letters, 26, 1964-1965, 1990.
[6] van Ooyen et al.: Improving the convergence of the back propagation algorithm. Neural Networks 5, 465-471, 1992.
[7] Shoemaker et al.: Back propagation learning with trinary quantization of weight updates. Neural Networks, 4, 231-241, 1991.
[8] Weigend et al.: Back propagation, weight elimination and time series prediction. In: Touretzky, D.S. Elman. J.L. Sejnowski T.J. Hinton. G.E.(Eds). Proceedings of the 1990 Connectionist Models Summer School. CA. 105-116, 1990.
[9] Wang et al.: The unpredictability of standard back propagation neural networks in classification applications. Management Science, 41(3), 555-559, 1995.
[10] Lee et al.: Improvement of function approximation capability of back propagation neural networks. Proceedings of the International Joint Conference on Neural Computation. 1, 541-551, 1991.
[11] LeCun et al.: Back propagation applied to handwritten zip code recognition. Neural computation, 1, 541-551, 1989.
[12] Hsiung JT et al.: Should back propagation be replaced by more effective optimization algorithms. Proceedings of the International Joint Conference on Neural Networks, (IJCNN), 7,353-356, 1990.
[13] Matsuoka et al.: Back propagation based on the logarithmic error function and eliminating of local minima. Proceedings of the International Joint Conference of Neural Networks. Singapore, 2, 1117-1122, 1991.
[14] Plaut et al.: Experiments on learning by back propagation (CMU-CS-86-126). Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA, Prentice Hall, India, 1998.
[15] Werbos P Jet al.: Back propagation thought time: What it does and How to do it. IEEE, 1550-1560, 1990.
[16] White H et al.: Some asymptotic results for back propagation. Proceedings of the IEEE Conference on Neural Networks, 3.261-66, 1987.