Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE) Vol. 2, Issue 2 (2015) ISSN 2349-1469 EISSN 2349-1477

An Efficient Method to Optimize Multi-Layer Perceptron for Classification of Human Activities

Murat Taşkıran, Zehra Gülru Çam and Nihan Kahraman

***Abstract*—**MLP which is one of the most commonly used classifier, is a feed-forward, supervised neural network topology. Back propagation algorithm is used for minimizing the error between network output and the target value. According to classification process, MLP structure and learning parameters, which are used in back propagation algorithm, are needed to decide for increasing the test accuracy. Commonly these variables are chosen randomly, so finding the values that give maximum test accuracy is a time-consuming process. In this paper, learning parameter of back propagation algorithm and network structure are optimized to success a faster and efficient weight-update process by using three different heuristic optimization algorithm, ABC, GA and SA. Both of the used two datasets contain human activity sensor data. For two datasets, three algorithms are compared and detailed test results are given. It is observed that, although SA is the fastest one among chosen algorithms, ABC shows the highest performance for test accuracy of MLP classifier.

***Keywords*—** multi-layer perceptron, artificial bee colony, genetic algorithm, simulated annealing algorithm, human activities classification.

I. Introduction

THE main drawback of the neural networks is the decision of the parameters via trial and error, or approximately. If learning rate in the Back Propagation Algorithm is chosen bigger, learning becomes faster but the risk of oscillation appears. But, if learning rate is small, learning process takes very long time. Same situation is also observed in Momentum Coefficient. Hidden neuron number is determined totally random. Selection of different activation function for each problem solution can give better solution according to output characteristic. Because of reaching the decision of all these values is time consuming, the development of an optimization algorithm is necessary.

There are various studies in the literature with Artificial Bee Colony(ABC),Simulated Annealing Algorithm(SA) and Genetic Algorithm(GA) regarding the cases, in which one parameter is optimized while all the others are held constant[1,2].There are limited number of studies which optimizes more than one parameter by using one algorithm[3]. In addition to these studies, Çam, Çimen and Yıldırım's work optimizes 3 parameters, hidden neuron number, learning rate and momentum coefficient, of back propagation algorithm[4]. In this study, we will be adding type of the activation function as the fourth parameter. The organization of this paper is as follows: The second section contains a short review of multi-layer perceptron and back propagation. The third section focus on the optimization algorithms used in this study. In the fourth section, the datasets are introduced and optimization process is explained. Finally in the last section, results are given.

Murat Taşkıran, Zehra Gülru Çam, and Nihan Kahraman, are with Yıldız Technical University Department of Electronics and Communication Engineering

II. Multi-Layer Perceptron

Multi-Layer Perceptron(MLP) is an essential tool for solving classification, identification and generalization problems. A basic three layer MLP topology is given in figure 1.

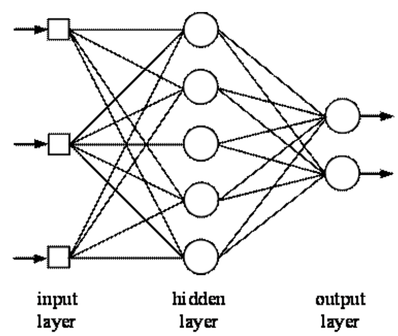

Fig. 1: A basic topology of MLP

Each neuron in the first layer, takes features of each sample, generates a weighted summation and gives this summation to an activation function as the function variable. Outputs of activation functions are outputs of the neurons. According to problem, a suitable activation function is chosen.

Number of input neurons are determined with input feature vector size and number of output neurons are determined according to number of classes. However number of hidden layer and neuron number of hidden layers are chosen generally with trial and error methods by designer. This artificial neural network model is using back propagation algorithm to minimize the squared error between networks output and target values. Learning process updates weights in each iteration with equation 1.

$$\Delta W_{ij}(t) = -\alpha dE / dW_{ij} + \beta \Delta W_{ij}(t-1)$$
(1.a)

$$W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij}(t)$$
(1.b)

In these equations, E is the error between network output and the target value, w is the weight between jth and ith neurons, α is learning rate and β is the momentum coefficient. Learning rate and momentum coefficient, which are used in back propagation algorithm, are chosen with the same method.

III. Optimization Algorithms

In this study, Artificial Bee Colony, Genetic Algorithm and Simulated Annealing Algorithm are used as optimization algorithms.

A. Artificial Bee Colony

Artificial bee colony algorithm is developed by modeling the honey searching behavior of bees. It is designed by Karaboğa in 2005 [5]. In this algorithm, each of the food source represents a solution for the problem. The bees in the colony search for the source with more nectar amount. Nectar amount corresponds to the fitness value of the food source.
The algorithm is divided into four steps:

- In Employed Bee Phase, employed bees search for food source randomly. At the first step of the algorithm, these sources are generated randomly by using the equation.

$x_i = l_i * \delta\ (u_i - l_i)$

In this equation, δ is a random number in the range [0,1]. l_i is the lower limit and ui is the upper limit.

- In Onlooker Bees Phase, bees calculate nectar amount of each randomly selected source and they look for the new sources with more nectar in the neighbourhood of the each source.

$$x_y = x_i + (x_i - x_k)$$

For each loop, the sources with the highest nectar amount are memorized.

- In Scout Bees Phase, the source is abandoned after the number of limit neighborhood search if the nectar amount is unable to be enhanced.
- Memorize the best solution

B. Genetic Algorithm

Genetic algorithm is the most common evolutionary optimization and search method. The algorithm has four main steps: Selection, Reproduction, Crossover and Mutations. Selection selects the individuals, called parents, that contribute to the population at the next generation. Crossover combines two parents to form children for the next generation. Mutation applies random changes to individual parents to form children. The flow chart of the genetic algorithm is given as below.

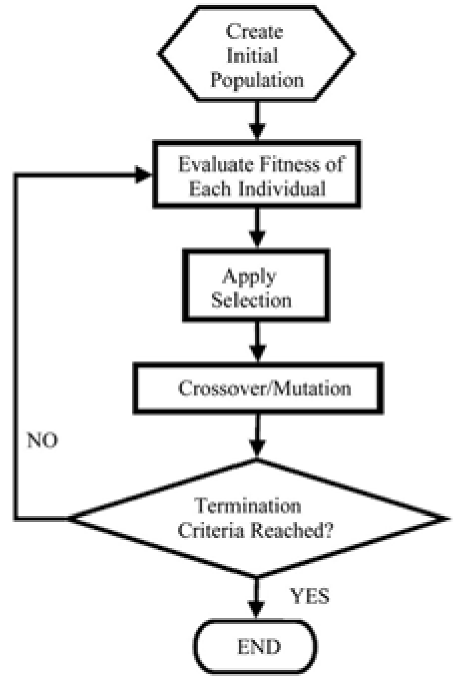

Fig. 2: Typical flowchart illustrating the steps of genetic algorithm

In genetic algorithm, the most important factor for evaluating the success of the problem solving is representation of the individuals which symbolize solutions of the problem[6]. There is a fitness function to decide if an individual is a solution for the problem or not. If the fitness value of the individual is high enough, this individual has chance to reproduce with other individuals in the population. These individuals produce new individuals with mutation. This new individuals have the characteristics of their parents. Because when new individuals are produced, individuals have low fitness value are selected rarely, after a little while these individuals are excluded from the population. New generation is formed by individuals with high fitness value and new generation includes most of the characteristics from the former generation. Thus, along plenty of generations good characteristics spread in the population and through the genetic operations, they combine with other preferred characteristics [7].

C. Simulated Annealing Algorithm

The simulated annealing algorithm is developed by behaviour of physical systems by melting a substance and lowering its temperature slowly until it reaches the freezing point (physical annealing).

SA decreases temperature slowly and in each temperature considers a neighbour of current state. If energy of neighbour is better than the current state, SA moves the system to neighbour state but if energy of neighbour was lower than the current state, SA uses an acceptance function to decide whether to move to neighbour state or not[8].

The procedure is controlled by a group of parameters which are called cooling schedule, so it can give near optimum to the problem in a reasonable time. The algorithm accepts new solution with the probability pt :

$$P_t(\alpha \Rightarrow \alpha') = \begin{cases} 1 & \text{while} f(a') \le f(a) \\ \exp(\frac{f(\alpha) - f(\alpha')}{t}) & \text{otherwise} \end{cases} \quad (2)$$

In formula (2), f is the objective function, α is the current solution, α' is a new solution, t is the current temperature[9].

IV. Optimization Process

The first dataset of the two used in the research Smartphone-Based Recognition of Human Activities and Postural Transitions dataset from UC Irvine Machine Learning Repository dataset contains 10929 sample with 561 feature. Activity recognition dataset built from the recordings of 30 subjects performing basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors. In the research, 6 classes, walking, walking upstairs, walking downstairs, sitting, standing and laying, are used. The 6 classes contain approximately 95% of the dataset. %65 of the dataset is used for training and the rest is used for testing.

The second dataset contains 1500 samples with 45 features. These features are the xyz coordinate values for hand, waist and knee in period of 1 second interval in a five second activity. The number of the output classes is 5: falling, standing, jumping, walking and sitting. This dataset has been collected for the research conducted in Yildiz Technical University by the year 2014[10]. 1250 out of this 1500 are used for training and the rest is used for testing.

For network topology, a multi-layer perceptron with one hidden layer is designed. Network to train with using back propagation algorithm is the target. Initial weights are selected randomly. 100 iterations are implemented for training. When test accuracy is calculated, for same learning rate, momentum coefficient, hidden neuron number and activation function type, test accuracy is calculated for 5 times with randomly selected initial weights, and the arithmetic mean of these 5 values are calculated. The one hidden layered MLP network formed to classify these datasets is optimized respectively with ABC, GA and SA. Parameters aimed to be optimized are hidden neuron numbers, learning rate, momentum coefficient and activation function type. Available options for activation function are sigmoid, bipolar sigmoid, linear and step function. For ABC algorithm, optimization is applied by the colonies with different volumes. The limit factor is chosen as 5. The collected nectar amount corresponds to the test performance. For GA, optimization is applied by different population sizes and generations.

V. Conclusion

The results classify HAPT dataset are shown in the Table I,II and III respectively by using ABC, GA and SA.

TABLE I: HAPT dataset with ABC Algorithm

# of the Bees	Iteration	Learning Rate	Momentum Coefficient	Hidden Neuron Number	Activation Function	Test Performance
6	10	0.4474	0.6329	111	linear	95.5%
8	8	0.2909	0.0316	596	sigmoid	93.67%
10	5	0.4111	0.4966	309	sigmoid	93.5%
12	3	0.4004	0.7487	71	linear	95.33%
14	2	0.7241	0.312	46	linear	95.33%

TABLE II: HAPT dataset with GA

Population	Generation	Learning Rate	Momentum Coefficient	Hidden Neuron Number	Activation Function	Test Performance
5	10	0.4295	0.5805	12	sigmoid	88%
10	10	0.2199	0.6598	112	linear	90.5%
10	15	0.4268	0.5321	151	Bipolar sigmoid	91.5%
15	15	0.1012	0.6410	458	Bipolar sigmoid	92.16%
15	20	0.411	0.4966	309	sigmoid	93.5%

TABLE III: HAPT dataset with SA Algorithm

Iteration	Learning Rate	Momentum Coefficient	Hidden Neuron Number	Activation Function	Test Performance
5	0.5	0.5	200	linear	90.67%
10	0.515	0.417	265	linear	92%
15	0.533	0.429	252	linear	92.16%
20	0.889	0.026	22	Bipolar sigmoid	93.33%
40	0.646	0.354	208	sigmoid	93.66%

The results classify second dataset are shown in the Table IV,V and VI respectively by using ABC, GA and SA.

TABLE IV: SECOND DATASET WITH ABC ALGORITHM

# of the Bees	Iteration	Learning Rate	Momentum Coefficient	Hidden Neuron Number	Activation Function	Test Performance
6	12	0.441	0.2932	101	step	86.6%
8	9	0.0117	0.4553	390	step	87.4%
10	8	0.0074	0.2482	237	step	87.2%
12	6	0.0052	0.6278	5	Bipolar sigmoid	86.4%
14	5	0.0074	0.2482	237	step	87.2%

TABLE V: SECOND DATASET WITH GA

Population	Generation	Learning Rate	Momentum Coefficient	Hidden Neuron Number	Activation Function	Test Performance
5	10	0.4604	0.7091	187	linear	74.6%
10	10	0.6280	0.4624	351	Bipolar sigmoid	75.4%
10	15	0.0011	0.7892	320	Bipolar sigmoid	78.2%
15	15	0.0019	0.6476	265	Bipolar sigmoid	81.8%
15	20	0.0052	0.6278	5	Bipolar sigmoid	81.4%

TABLE VI: SECOND DATASET WITH SA ALGORITHM

Iteration	Learning Rate	Momentum Coefficient	Hidden Neuron Number	Activation Function	Test Performance
5	0.549	0.549	58	linear	73.04%
10	0.164	0.847	223	Bipolar sigmoid	75.18%
15	0.011	0.043	32	linear	77.2%
20	0.974	0.06	4	Bipolar sigmoid	78.6%
40	0.909	0.081	324	Bipolar sigmoid	78.8%

For ABC algorithm, iteration has continued until the highest success rate is reached in colonies containing different number of bees. At the end, it is seen that in colonies with higher number of bees; highest success rate is reached with less iteration. Maximum number of iteration has been selected for stopping criteria in GA algorithm. Optimization processes has been initiated for different population values. In SA algorithm, initial temperature has taken as 100. Again here, the stopping criteria is selected as the number of iterations. At the end, the GA algorithm has been the one that gave the slowest amongst all others. SA algorithm, on the other hand, has performed well in terms of speed but could not obtain as successful results as the others. Therefore, the ABC algorithm has been the most successful one in terms of obtaining best results.

ACKNOWLEDGMENT

This material is based upon work supported by the Ministry of Science, Industry and Technology, Republic of Turkey project San-Tez 0817.STZ.2014.

REFERENCES

[1] T. B. Ludermir, A. Yamazaki, and C. Zanchettin, "An Optimization Methodology for Neural Network Weights and Architectures", Neural Networks, IEEE Transactions on vol. 17 , issue.6 pp.1452 – 1459,November 2006 http://dx.doi.org/10.1109/TNN.2006.881047

[2] M. J. Er and F. Liu "Genetic Algorithms for MLP Neural Network Parameters Optimization", presented at the IEEE Control and Decision Conference, Chinese, 2009.

[3] M. Watts, L. Majorb and W. Tateb, "Evolutionary Optimisation of MLP for Modelling Protein Synthesis Termination Signal Efficiency", presented at the IEEE Congress on Evolutionary Computation, May 2002. http://dx.doi.org/10.1109/cec.2002.1006232

[4] Z.G. Çam, S. Çimen and T. Yıldırım "Learning parameter optimization of Multi-Layer Perceptron using Artificial Bee Colony, Genetic Algorithm and Particle Swarm Optimization", presented at IEEE International Symposium on Applied Machine Intelligence and Informatics(SAMI), Slovakia 2015.

[5] D. Karaboga, "An Idea Based On Honey Bee Swarm For Numerical Optimization", Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department 2005.

[6] C. Juang "A Hybrid of Genetic Algorithm and Particle Swarm Optimization for Recurrent Network Design", IEEE Trans. on Systems, Man, and Cybernetics. vol.34 pp. 997-1006, April 2004 http://dx.doi.org/10.1109/TSMCB.2003.818557

[7] D. W. Boeringer and D. H. Werner "Particle Swarm Optimization Versus Genetic Algorithms for Phased Array Synthesis", IEEE Trans on Antennas and Propagation, vol. 52 pp. 771-779, March 2004. http://dx.doi.org/10.1109/TAP.2004.825102

[8] M. M. Keikha, "Improved Simulated Annealing using Momentum Terms", presented at the IEEE Second International Conference on Intelligent Systems, Modelling and Simulation, Kuala Lumpur, Jan 25-27,2011. http://dx.doi.org/10.1109/isms.2011.18

[9] Q. Ji-Yang "Application of Improved Simulated Annealing Algorithm in Facility Layout Design" presented at the IEEE Control Conference(CCC), Beijing July 29-31,2010.

[10] O. Kurban, "İnsan Aktivitelerinin Giyilebilir Sensörler ile Özellik Çıkarımı Yapılmadan Sınıflandırılması", M.S. thesis, Dept. Electronic Eng., Yildiz Tech. Univ., İstanbul, Turkey, 2014.