Cost Optimization in Neural Network using Whale Swarm Algorithm with Batched Gradient Descent Optimizer

C Karthikeyan¹, E Sreedevi¹, Naveen Kumar¹, E Vamsidhar¹, T Rajesh Kumar¹, and D Vijendra Babu²

¹Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur, Andhra Pradesh, India
²Department of Electronics and Communication Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission’s Research Foundation, Chennai, Tamil Nadu, India
E-mail: ckarthik2k@gmail.com

Abstract. Optimization algorithms are liable for sinking the losses and to give the most precise outcomes conceivable. Optimizers are utilized to modify the properties of neural network, for example, training rate and weights are used to reduce the losses. Optimization means a procedure of obtaining a global optimal solution for a given problem under given conditions. The real-world problems in the scientific fields, such as engineering design and economic planning, mostly are multimodal, high-dimensional, disconnected, and oscillated optimization problems. These complex problems cannot be solved well within reasonable time using traditional method based on gradient. Nature-inspired algorithms are becoming delightful in resolving mathematical optimization problems, like multiprocessor scheduling problem, vehicle routing and classification problems etc. In this manuscript, Whale Swarm Optimization algorithm on optimizing the neural networks, one of the meta-heuristic algorithms is applied to analysis of the cardiovascular disease dataset and compares the performance with Gradient Descent and RMSprop optimization techniques.

Keywords: Whale Swarm Optimization, Gradient Descent, Neural Network

1. Introduction

The network of neural is developed using human brain’s architecture with neurons. The neurons in the neural network have the functionality of taking the input and generate an output by applying the input on a function. These functions are entitled as activation functions [1]. The neural network carries three layers namely output, input and hidden layers. Each layer in the neural network holds a bunch of neurons; the number of features chosen is given with number of neurons in input layers. The aggregate number of hidden layers depends upon the model, data size and the complication in the problem [2]. The
hidden layer’s neuron count does not depend on the number of features, and the logistic functions to observe the probability score of each class at the output layer. By the application of biases or weights in the activation function and hidden layer gives a nonlinear network [3]. The accurate prediction of weights to be applied at the hidden layer can be predicted by back propagation algorithm. In Genetic Algorithm, the performances are significantly affected, when two new individuals are crossover to develop a new one of a genetic system. A permutation encoding for crossover operators are described in reference [4]. A random diversity in the population is provided with another operator called as Mutation in Genetic Algorithm [5], to avoid premature achievement of algorithm. Michalewicz had suggested uniform-turbulence mutation and non-uniform mutation [6] for optimization problems in numerical. Deb also suggested that most broadly used operator is polynomial mutation [7]. The manuscript [8] deals with comprehensive introduction to mutation operator. When working with different optimization problem, it is very important to design or select appropriate crossover and mutation operators of Genetic Algorithm. For minimizing nonlinear and non-differentiable continuous space functions, author developed a differential evolution algorithm in [9].

The real world applications in optimization based on particle swarm optimization (PSO) algorithm resolve many difficult and complex problems [10-12]. Based on renewal of value in position and velocity, every particle moves towards a new position, in which the cognitive better position is based on velocity is generally discussed in traditional PSO algorithm [13]. So far, different optimization problems are solved by many types of PSO variants. For example, a balanced local and global search are being introduced by Shi et al. for linear decreasing inertia weight into PSO (PSO-LDIW)[14]. In function optimization, the auto control of parameter improves the efficiency of convergence speed and search, in which, a strategy of learning is a local optima jumped out is discussed by [15] Zhan et al. in Adaptive PSO (PSAO).

2. Whales Optimization Algorithm

A novel Whale Optimization Algorithm (WOA) is a population-based meta-heuristic algorithm proposed by Nasiri et al., [16] which is stirred by the hunting behaviour of humpback whales. It has the advantages of few control parameters and simple structure. However, we find that a local optimum is easy when having relation with high-dimensional complex optimization problems. The quality of solution obtained by WOA needs to be further improved.

![Figure 1 Optimization Performance of Meta-Heuristic Algorithms](image)

The investigators have identified that the population diversity and evolutionary direction play an important role in the optimization performance of meta-heuristic algorithms. However, the population diversity shrinks rapidly in the later stage of optimization process.
leading WOA easily falling into local optimum.

3. Whales Hunting Mechanism

In ocean, more than 80 whale species are available with great physical capacities and intellectual aquatic mammals. The whale’s acoustic effect is linked the function such as feeding, migration and mating patterns, since it has wide acoustic range and lives in group. The ultrasound of the whale is not human audible. The quantity and quality of the prey of a whale is informed to other whales, when a whale found its food. Thus neighbor whales inform the main whales by sending signals, and migrate to certain position to attain the food. In this way, the whales corresponds one another and it leads to hunt its pray, which develop an optimization problem algorithm (i.e) meta-heuristic algorithm.

![Figure 2 Whales Hunting Mechanism](image)

Hunting rules of whale are utilized for describing Whale Swarm Algorithm [17]. The following four important protocols are used:

1. In searching area, the ultrasound is used by all the whales for communicating each others.
2. The computing capability is available for each whale to calculate the distance of another one.
3. The fitness is associated to all whales at a rate of finding a quantity and quality of food.
4. The better judgment by fitness is obtained by one among the other whale guided each other in movements.

The electromagnetic waves not require any medium of propagation, generated by light and radio signals. Nevertheless a medium is requiring for acoustic waves to travel, such as air, wood, metal and water. The Ultrasound signal comes under the group of acoustic wave, its speed of transmission and distance of transmission deal with its medium. From a source at a distance ‘d’, the ultrasound having ‘ρ’ as its intensity, its origin of source intensity is ‘ρ₀’ with natural constant e, then,

\[ ρ = ρ₀ \cdot e^{-\eta d} \]

Whereas, the attenuation coefficient is ‘η’, deals with the attributes of medium and ultrasound. The objective functions, its dimension, distribution of peaks, range of variables are getting affected in a factor of ‘η’. So, ‘η’ value is carefully set for various objective functions, since the ultrasound travelling distance is too long. The understanding power of one whale to another is not sure, if the distance is very long. Thus, the movement of a whale is identified as best, when its random and negative movements are calculated. So, after some time, a whale swarms forms, with best neighbor whale having random movements as its feature.
4. Whale Swarm Optimization (WSO) with Batched Gradient Descent Optimizer

Implementing WSO is a difficult task in neural network, so we have to introduce the batched gradient descent technique idea with whale swarm optimization methodology. The Characteristics of WSO often preferred in numerical search optimization better known for identification of hyper parameters but not for learning. However, we can build a model where we can solve the new learner’s objective function with whale swarm optimizer that was produced by another learning optimizer.

The needs to focus on choosing the learning algorithm are
i) Compatibility: This algorithm should be compatible with WSO. Since WSO searches for global maxima in groups, the algorithm should be able to understand the groups. One of the ways is through continuous evaluation.
ii) Direction: WSO’s grouping technique should be guided by this algorithm. This means the global leader is the one appointed by this optimizer.
iii) Continuity: Learning should be continuous, and WSO should not refresh itself for very iteration.

Based on these parameters Batched Gradient Descent optimizer (BGD) is chosen to best suitable with WSO. WSO is suitable for the neural network optimization for cost optimization and as cost plays a important role in neural networks. The combination of both the BGD along with the WSO algorithm makes it even better to predictions, and majorly helps in optimizing the neural networks.

4.1 Algorithm

Using the above protocol, the proposed Whale Swarm Optimization Algorithm having a general framework as:

Input: The whale swarm Ω is the objective function.
Output: The optimized global output.
Step 1: begin
Step 2: Initialize parameters.
Step 3: Initialize whales’ positions;
Step 4: All the whales are evaluated, in which their fitness is calculated.
Step 5: while termination criterion is not satisfied do
Step 6: for i=1 to |Ω| do
Step 7: Find the nearest and best whale Y of Ωi;
Step 8: if Y exists then
Step 9: According to equation 2, Ωi moves under the guidance of Y;
Step 10: Evaluate Ωi;
Step 11: end if
Step 12: end for
Step 13: end while
Step 14: return the global optima.
Step 15: end

With the Initialized parameters, we start and the objective function later we also initialize the positions of the whale at which they are located. Now, randomly assign the entire whale to the search area. Then each whale gets evaluated for its suitable and the matching whales are taken into consideration are fragile than the other whales, while comparing. This process will continue until the termination criteria or the fittest of the population of whales is found. Later, the number of members in Ω is indicated by |Ω| as per step 6, specifically the Ω and swarm size in step 7 is the i\textsuperscript{th} whale in Ω. From those steps, the initial steps are computation of iterative, consist of initializing positions, configuration parameters and calculating individually, which are alike other meta-heuristic algorithms.
Now, in search area, randomly all the whales are step in. Hereafter, as a core of WSA: whales movement is from step 5-13. All whales move towards the good prey with full cooperation. Next, a whale should find its better and nearest whale (step 7).

5. Gradient Descent

A vector-valued function is gradient that represents the slope of the tangent of the graph of the function, pointing the direction of the greatest rate of increase of the function. It is a derivative that indicates the incline or the cost function slop. A common optimization algorithm is Gradient Descent in machine learning and deep learning. It is a first-order optimization algorithm. This means it only considers the first derivative when performing the updates on the parameters.

On each iteration, we update the parameters in the reverse path of the gradient of the objective function $J(w)$ with respect to the parameters, in which gradient provides the direction of the steepest ascent. The size of the steps are registered on each iteration to reach the local minimum is determined by the learning rate $\alpha$. Therefore, we follow the direction of the slope downhill until we achieve a minimum value. The gradient is a vector-valued function, and as a vector, it has both a path and a magnitude. The algorithm of Gradient descent multiplies the gradient by a number (Learning rate or Step size) to determine the next point.

The learning rate is nothing but the step size. We can cover more ground step with high learning rate, but it’s overshooting the lowest point, since there is a constant change in hill slope.

![Figure 3 Gradient Descent Learning Rate](image)

With an extraordinary frequent recalculation, the low learning rate moves confidently towards the negative gradient. It is a time consuming process to calculate gradient, because of slow learning rate and it is more accurate, so it takes more time to settle down. As the gradient have the relationship with cost function, it has its own curve. In order to get an accurate model, the parameters are updating its value and the curve shows the slop.

Algorithm:

Step-1: Initialize a very small random value to weights.
Step-2: For each weight $w_{ij}$ set
\[ \Delta w_{ij} := 0 \]
Step-3: For every data point $(x, t)_p$
1. set input units to $x$
2. compute value of output units
3. For each weight $w_{ij}$ set
\[ \Delta w_{ij} := \Delta w_{ij} + (t_i - y_i) y_j \]
Step-4: Repeat step 2-3 until done
Step-5: Now $w_0$ is set for each weight.
6. RMSprop Optimizer

The RMSprop optimizer is like the gradient descent algorithm with momentum. The RMSprop optimizer restricts the oscillations in the vertical direction. Therefore, learning rate may be increased and our algorithm could take larger steps in the horizontal direction converging faster. How the gradients are calculated shows the difference between RMSprop and gradient descent.

Consider an example where we are trying to optimize a cost function which has contours like below and the position of the local optima (minimum) is indicated by red dot.

![RMSprop Learning Rate](image)

**Figure 4 RMSprop Learning Rate**

We start gradient descent from point ‘A’ and after one iteration of gradient descent we may end up at point ‘B’, the other side of the ellipse. Then another step of gradient descent may end up at point ‘C’. With each iteration of gradient descent, we move towards the local optima with up and down oscillations. If we use larger learning rate then the vertical oscillation will have higher magnitude. So, this vertical oscillation slows down our gradient descent and prevents us from using a much larger learning rate.

7. Results and Discussion

The cardiovascular dataset [18] contains the patient’s information in the form of columns and those are id; age; gender; height; weight; cholesterol; glucose; active; bp; cardio. From the information in the dataset the predictions are done regarding the patient’s heart condition. To know such condition, his/ her bp levels and cholesterol levels are more important. These two components help in analysis of the patient’s condition and prediction of cardiac disease may likely be affected or not. The patient’s habits resemble the treatment to be given. We can use such data for different analysis.

![Costs and Accuracy for RMSprop Optimizer](image)

**Figure 5 Costs and Accuracy for RMSprop Optimizer**
The Figure 5 shows the Cost and Accuracy for RMSprop Optimizer. The output result denotes that for the RMSprop algorithm the cost and accuracy are too high or too low, respectively. The Figure 6 shows the Cost and Accuracy for Whale Swarm Optimization. It shows that, the whale swam algorithm reached peak cost value but decreases instantly and when it comes to accuracy it’s constantly dataset fluctuating and observed mostly to be low. These fluctuations are common as we use a real-world.

![Figure 6 Costs and Accuracy for WSO](image)

The Figure 7 shows the Cost and Accuracy for Gradient Descent. Here, accuracy is directly proportional to cost. When there is a change in cost, then there is an appropriate change in the accuracy.

![Figure 7 Cost and Accuracy for GD](image)

The Figure 8 shows the Comparison of Cost Function Value of Optimizers. The entire three optimization algorithms are compared; they are gradient descent, Whale Swarm and RMSprop. The comparison involved the cost function between these three algorithms and we can see that when we compare them the Whale optimization reached peak cost value but decreases instantly comparing with other optimizers WSO show lesser cost.

![Figure 8 Comparison of Optimizers](image)
In Figure 9, it shows the Comparison of accurate value of optimizers. It contains a detailed comparison of accuracy values between all three optimization algorithms, gradient descent, Whale Swarm and RMSprop. When we consider accuracy, WSO may not be the best accurate model but it reached peak accuracy multiple times which means, with some external help, we can build a highly accurate model using WSO comparing to other algorithms.

8. Conclusion

When we consider the cost, whale swam is the best, but its lacking behind when we consider the accuracy between these three optimization algorithms Gradient Descent, RMSprop and Whale Swarm, however, WSO is the suitable fit for neural network optimization for cost optimization and as cost plays a crucial role in neural networks. Accuracy is comparatively low for WSO algorithm but still, it’s not much suitable when accuracy is given highest preference while analysis of the real world data which gives a future scope for improving this algorithm to perform better.
References

[1] Cheerla, S., & Ratnam, D. V. 2018 RSS based wi-fi positioning method using multi layer neural networks, Spaces, 58-61. doi:10.1109/SPACES.2018.8316316.

[2] Pellakuri, V., & Rajeswara Rao, D. 2016 Training and development of artificial neural network models: Single layer feedforward and multi layer feedforward neural network, Journal of Theoretical and Applied Information Technology, 84(2), 150-156.

[3] Karthikeyan, C., Priyanka, G.B.N.D., Jeewana, K., Sai, M.K. 2020 Classification of white blood cells using convolutional neural networks, International Journal of Advanced Science and Technology, Vol.29, No.7 (2020) pp.886-892.

[4] T. Kellegöz et al 2008 Comparing efficiencies of genetic crossover operators for one machine total weighted tardiness problem, Applied Mathematics and Computation, 199 590-598.

[5] Patle, B.K et al 2018 Matrix-Binary codes based Genetic Algorithm for path planning of mobile robot, Computers & Electrical Engineering DOI:10.1016/j.compeleceng.2017.12.011.

[6] Z. Michalewicz, 2013 Genetic algorithms+ data structures= evolution programs, Springer Science & Business Media.

[7] Reddy et al 2018 An effective hybrid multi objective evolutionary algorithm for solving real time event in flexible job shop scheduling problem. Measurement.

[8] Prabhu et al 2019 EODVGA: An Enhanced ODV Based Genetic Algorithm for Multi-Depot Vehicle Routing Problem, EAI Endorsed Transactions on Scalable Information Systems, 10-6.

[9] R. Storn, K. Price, 1995 Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces, ICSI Berkeley.

[10] Raju, N. R., & Reddy, P. L. 2016 Robustness study of fractional order PID controller optimized by particle swarm optimization in AVR system. International Journal of Electrical and Computer Engineering, 6(8), 2033-2040. doi:10.11591/ijece.v6i8.10633

[11] Kumar, T. R., & Gattu, K. R. 2017 Multi objective particle swarm optimisation (PSO) algorithm for ideal siting and sizing of distributed generation (DG) systems. Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 18), 832-844.

[12] Prasad, C. M. V. S., Rao, K. R., Satish Kumar, D., & Prabhu, A. V. 2018 Performance evaluation of power optimization in wireless sensor networks using particle swarm optimization. International Journal of Engineering and Technology (UAE), 7, 404-408.

[13] Jyothula, H., Rao, S. K., & Vallikumari, V. 2017 Two phase active counter mechanism embedded with particle swarm optimization technique for segmentation of bio-medical images. Journal of Advanced Research in Dynamical and Control Systems, 9(6), 232-242.

[14] Y. Shi, R. Eberhart, 1998 A modified particle swarm optimizer, in: Evolutionary Computation Proceedings, IEEE World Congress on Computational Intelligence., The IEEE International Conference on, IEEE, pp. 69-73.

[15] Z.-H. Zhan, J. Zhang, Y. Li, H.S.-H. Chung, 2009 Adaptive particle swarm optimization, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 39 1362-1381.

[16] J. Nasiri and F. M. Khiyabani, 2018 ‘A whale optimization algorithm (WOA) approach for clustering’, Cogent Math. Stat., vol. 5, no. 1.

[17] Rajesh Kumar, T., Suresh, G.R., Kanaga Subaraja, S. & Karthikeyan, C. 2020 “Taylor-AMS features and deep convolutional neural network for converting nonaudible murmur to normal speech”. Computational Intelligence, 20 pp. 1-12.

[18] Rajesh, N., Manesha, T., Hafeez, S., & Krishna, H. 2018 Prediction of heart disease using machine learning algorithms. International Journal of Engineering and Technology (UAE), 7(2.32 Special Issue 32), 363-366.