Multi-objective Cuckoo Search Algorithm based training of Convolutional Neural Network

Moid Ul Huda
Department of Computer Science
Habib University
Karachi, Pakistan
mh05205@st.habib.edu.pk

Muhammad Sabihul Hasan
Department of Computer Science
Habib University
Karachi, Pakistan
mh04387@st.habib.edu.pk

Syed Hammad Ali
Department of Computer Science
Habib University
Karachi, Pakistan
sa04324@st.habib.edu.pk

***Abstract*—Meta-heuristic algorithms are a class of algorithms that are used to solve optimization problems. These algorithms generate near optimal solution in limited resources and can be used to formulate variety of problems. Many of such algorithms are based on the naturally occurring phenomenon in the world and show how nature inspires us to find solutions that nature has developed over millenniums.**

In this paper We will be looking to formulate the problem of training a convolutional neural network on a new metaheuristic algorithm, called Cuckoo Search (CS), for solving optimization problems. Cuckoo search algorithm is a recently developed meta-heuristic optimization algorithm for solving optimization problems. This is a meta-heuristic algorithm inspired by nature. The algorithm is based on the parasitic nature of some cuckoo species along with Levy flights random walks.

***Index Terms*—Meta-heuristic, convolutional, cuckoo search, Lévy flight, optimization**

I. Introduction

The cuckoo optimization algorithm is based on the life of a bird called ”cuckoo”. The basis of this new optimization algorithm is the reproduction and specific spawning of the bird’s eggs. Adult cuckoo and eggs are used in this model. Adult cuckoos lay eggs in the habitat of other birds. If these eggs are not found and removed by the host bird, they will grow into mature cuckoos. We hope that the immigration and environmental regulations of the cuckoo colony can make them congregate and reach the best place for reproduction. The target function is this best place. Cuckoo optimization was developed by Yang and Deb in 2009, and it was inspired from the natural world. The Cuckoo optimization algorithm was developed by Rajabioun in 2011. The Cuckoo Optimization Algorithm (COA) is actually a new continuous search method based on the life of the cuckoo bird. Similar to other meta-heuristics, COA starts with a main population (a group of cuckoos). These cuckoos lay eggs in the habitat of other host birds. A set of potential solutions representing the habitat in the COA are generated.

II. Literature Review

In the paper [1], the authors discuss the drawbacks and limitations of traditional artificial neural networks and the back propagation algorithm that is used in it to learn and update the weights of the neurons in the structure. The authors have then proposed a new algorithm based on the nature inspired algorithm called Shuffled Frog Leaping Algorithm which is a memetic meta-heuristic algorithm combining the method of evolutionary algorithm with particle swarm optimization to produce a structure of ANN that has optimized synaptic weights for the purpose of pattern classification of images. The algorithm was tested with 10 benchmark datasets vs the back propagation algorithm and the results show that the SFLA performs better in most of the datasets against the back propagation algorithm.

In [2], the authors have described how different metaheuristic algorithms can be implemented to optimize the training process of a Wavelet neural network. They showed three very commonly used metaheuristic algorithms (Genetic Algorithm, Particle Swarm Optimization (PSO) and Harmony Search Algorithm) and compared their results for the validation of the ”No free food” rule for the optimization problems. The results shown in the article show that all the algorithms perform similar but the data set can probably give a slight advantage to some algorithms. But overall, metaheuristic algorithms optimize the training better than traditional methods.

The paper [3] talks about the contribution of Meta-heuristic Algorithms in optimizing feed forward artificial neural networks. Genetic Algorithm (GA) has been implemented by Leung et al to find the global optimum of the fitness function; thus, the parameter combinations of weight are the trained weight for the network. It is foreseen that classical meta-heuristics such as PSO and GA will be first applied to solve Deep Learning problems, followed by variants and hybrids of those.

In [4], the authors look into the details of feed-forward neural networks in detail starting with the history of the architecture, then looking into the details of components of FNN, influencing factors in FNN architecture, some common optimization techniques. In conclusion, FNN can be used for solving a wide variety of problems and meta-heuristic algorithms can be used to optimize them further and improve on the methods we currently use.

The paper [5] starts by discussing the advantages of using a meta-heuristic algorithm in optimizing the training process of a Convolutional Neural Network. It discusses the different meta-heuristic algorithms being applied to the deep learning algorithms to improve their performance in the literature being published. After testing the new optimised algorithm with MNIST and CIFAR10 datasets it showed that the new algorithm worked better in terms of performance while keeping the computational cost to same level.

In this paper [6], the authors have implemented three meta-heuristic algorithms on the training of convolutional neural network to optimize the weights for the layers of the CNN for better accuracy in classification problems. The paper has implemented and combined Simulated Annealing, Differential Evolution and Harmony Search with Convolutional Neural Network (LeNet-5) with 2 different variations in structure. The architectures are tested with 2 datasets, MNIST and CIFAR10 with both the variation in CNN structure and the results show that the structures with meta-heuristic optimization perform better than the usual CNN algorithms in all the cases whereas the time taken by each algorithm is also quite similar.

In this paper [7], a survey of training deep learning models with meta-heuristic algorithms is done. The neural network models described in the paper include the perceptron, feed-forward neural network and spiking neural network whereas the meta-heuristic algorithms mentioned are Genetic Algorithm and Particle Swarm Optimization with great detail.In the end, it shows that the underlying results in implementing meta-heuristic algorithm with neural networks/deep learning architecture show that it is a great area of research to spend resources into to optimize our models better.

In [8], the authors have proposed a method to optimize the process of training of Deep CNN architecture by multi-objective metaheuristic algorithm. The CNN architecture trained is DenseNet-121 which is one of the deepest convolutional neural networks out in the industry with the state-of-the-art classification accuracy and the algorithm used is called the OMOPSO which is a multi-objective Particle Swarm Optimization technique that uses particles working simultaneously to optimize the different parameters in the problem definition. The benchmark dataset that was used to train and test the architecture was the CIFAR-10 dataset. This showed that the multi-objective optimization works better than tradition training of the neural networks with less computational cost.

In [9], the authors have developed a new multi-objective metaheuristic algorithm based on the behaviour of sperm whales in nature. The algorithm is used to optimize the process of content based image retrieval and reduce the complexity of the model that would be used to classify the images according the the trained dataset. The algorithm was tested on 4 datasets for the accuracy of the results of the image retrieval as well image classification and the time taken for generating the results. A lot of testing results are shown in the paper which shows that the algorithm worked great on all the 4 datasets.

III. Cuckoo Breeding Behavior Strategy

A. Lévy Flights mechanism

Animals randomly look for food in the wild. The foraging path of the animal is actually a random walk, because the next action is based on the current position/state and the possibility of transitioning to the next position. The probability of the selected direction is mathematically modeled. A number of studies have shown that the flight behavior of many animals and insects show the typical characteristics of Lévy flights. Lévy flight is a random walk, in which the step length is calculated according to the heavy-tailed probability distribution. After a large number of steps, the distance to the starting point of the random walk tends to have a stable distribution.

B. Cuckoo Search Implementation steps

Cuckoo Search via Lévy Flights

begin
  Objective function $f(\mathbf{x}),\ \ \mathbf{x} = (x_1, ..., x_d)^T$
  Generate initial population of
    n *host nests* $\mathbf{x}_i\ (i = 1, 2, ..., n)$
  while *(t <MaxGeneration) or (stop criterion)*
    Get a cuckoo randomly by Lévy flights
      evaluate its quality/fitness F_i
    Choose a nest among n *(say,* j*) randomly*
    if $(F_i > F_j)$,
      replace j *by the new solution;*
    end
    A fraction (p_a*) of worse nests*
      are abandoned and new ones are built;
    Keep the best solutions
      (or nests with quality solutions);
    Rank the solutions and find the current best
  end while
  Postprocess results and visualization
end

Fig. 1. Pseudo code of the Cuckoo Search (CS).

Every egg in the nest is a solution, and the cuckoo egg is a newly occurred solution. The goal is to replace the not-so-good solutions with a new and possibly better solution (cuckoo). In its simplest form, each nest has one egg. The algorithm can be extended to more complex situations, in which each nest has multiple eggs representing a set of solutions (Yang 2009; Yang 2010). The CS algorithm is based on three idealized rules:

- Each cuckoo lays its eggs once and then throws it into a randomly selected nest.
- The best nest with high-quality eggs (solution) will be passed on to future generations.
- The number of host nests available is fixed, and the probability that the host can find foreign eggs with probability $p_a \in$ [0,1]. In this case, the host bird can throw the egg away or leave the nest and establish a brand new nest in a new location (Yang 2009).

Based on these three rules, the basic steps of the Cuckoo Search (CS) can be summarized as the pseudo code shown in Fig. 1.

For simplicity, the final assumption can be approximated by a fraction p_a of the n nests being replaced by new nests, having new random solutions. For maximization problems, the quality or applicability of the solution can be simply proportional to the objective function. Similarly, other forms of fitness can be defined by the fitness function in the genetic algorithm. Bacanin provides an object-oriented software implementation of cuckoo search. On the other hand, the unrestricted optimization problem defined by the modified cuckoo search algorithm. Other forms of fitness can be defined in a similar way to the matching function in genetic algorithms (Yang 2009).

IV. Multi Objective Cuckoo Search on CNN training

The algorithm that we present in this paper is a multi-objective cuckoo search algorithm that optimizes two objective functions that optimize two parameters of our deep neural network: maximize the the classification accuaracy of the network and minimize the FLOPs (floating point operations) that are carried out during the process of convolution. As the multi objective function optimizes two parameters separately, a Parato front output is generated which can be used to pick any model that suits the hardware of the user and can easily be used with the dataset.

A. Encoding of Deep CNN Parameters

The algorithm works by encoding the parameters of a Deep CNN as an individual in the population (nests in which the cuckoo bird lays the eggs). Specifically, the data is encoded into a vector of twice the size of number of deep blocks present in the network (for avoiding fluctuations in the model training, number of deep blocks are set to a constant). Every deep block has 2 parameters: number of layers in the block and growth rate of each block. These parameters will be used to calculate the objective functions hence they need to be optimized to get the produce the optimal results. The encoding vector for 3 deep blocks can be seen in figure 2.

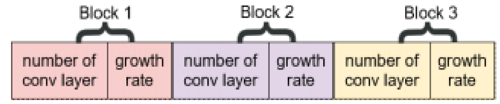

Fig. 2. Example of a vector encoding

B. Objective Functions for calculating fitness

The objective functions that are used to calculate the fitness of both the optimization parameters are based on the ResNet implementation of the Deep CNN. A model with similar architecture as the ResNet is constructed with the number of layers in each Deep Block coming from the encoding vector for each individual in the population. The deep blocks are separated with batch normalization and dropout layers for the training of the neural network. In the final layer Softmax activation function is used to validate the results of the neural network. Adam Optimization with the default parameters for learning rates and beta values are used ($\alpha = 0.001$, $\beta_1 = 0.9$. $\beta_2 = 0.999$). The main priority of the algorithm is to optimize the classification accuracy and the second priority is to minimize the FLOPs in the neural network.

These objective functions are used to define the fitness of an individual and perform the cuckoo search algorithm to filter out the less fit individuals from the population and insert new individuals. And the algorithms runs for some number of iterations or a termination criteria and in the end we get a list of models that are sorted on the basis of our objective functions with create a Pareto front for the best result to be selected.

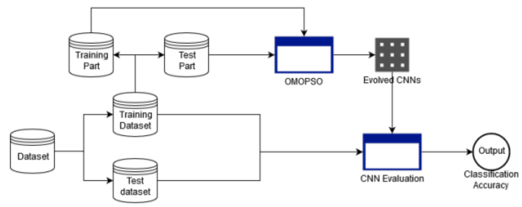

Fig. 3. Architecture of the Neural Network Presented

C. Overall training and working of the Neural Network Architecture

The architecture of the overall neural network works by first dividing the dataset into training and test samples and then the training set is further divided into 2 sets of training part and test part to train and test the evolution of network through the Cuckoo Search algorithm and to calculate the objective functions to optimize the parameters. As the metaheuristic algorithm results in a Pareto front, we choose the optimal network from the front for our use and test it on the actual test sample. The architecture can be seen in figure 3.

V. Experiments and Results

The CIFAR-10 dataset was used to train and test the multiobjective Cuckoo search based neural network. A sample image set can be seen in figure 4 Each individual in the population of the algorithm is a network similar to the ResNet architecture. The size of population is selected to be 50. The number of epochs, one individual was run for to calculate the accuracy metrics is 10 with a batch size of 128. The algorithm works on a number of iterations to optimize the parameters for the neural network. The iterations carried out in our experiment are 300.

As can be seen in the figure 5 and figure 6, the accuracy of the classification is on the rise and the loss is being reduced. Due to lack of resources on the GPU side, we were unable to completely run the algorithm for the complete iterations because as reported by [8], training of a very similar deep neural network took around 3 weeks on a single GPU unit. The results shown here show the progress of accuracy of the model.

Fig. 4. Sample images from the CIFAR-10 dataset

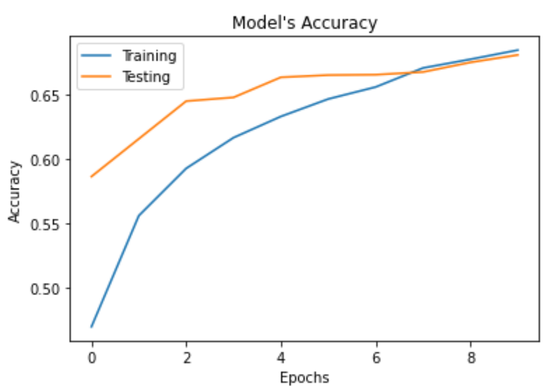

Fig. 5. Example of a vector encoding

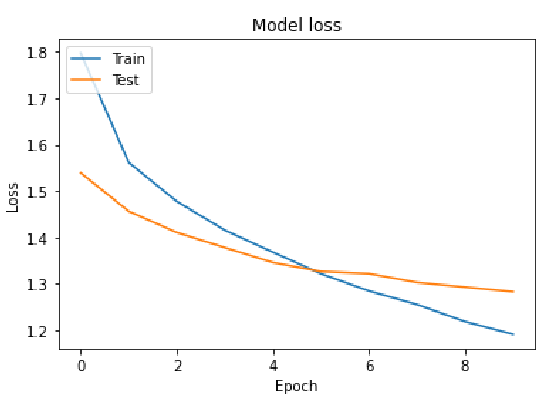

Fig. 6. Example of a vector encoding

VI. Conclusion

Meta-heuristic algorithms are a widely growing field of optimization techniques that can be used for a wide variety of optimization problems and perform better than most commonly used algorithm in less computational resources. Neural networks have been a widely researched in the field of computer vision but the optimization of the hyperparameters is a challenging problem that could be solved using multi objective optimizations. Our proposed algorithm fits into the category of multi objective metaheuristic algorithm that is designed to optimize the hyperparameters of a deep neural networks which take a lot of time and resources for their training and testing and due to a lot of parameters, optimizing the networks is difficult and hence can be solved with meta heuristic algorithms such as Cuckoo Search Algorithm.

Acknowledgment

The authors would like to thank Dr. Muhammad Farhan for useful discussions and suggestions and for giving the authors the opportunity to work on this research.

References

[1] Shaimaa Alwaisi. Training of artificial neural network using metaheuristic algorithm. *International Journal of Intelligent Systems and Applications in Engineering*, SpecialIssue:12–16, 07 2017.

[2] Kee Huong Lai, Zarita Zainuddin, and Pauline Ong. A study on the performance comparison of metaheuristic algorithms on the learning of neural networks. volume 1870, page 040039, 08 2017.

[3] Simon Fong, Suash Deb, and Xin-she Yang. How meta-heuristic algorithms contribute to deep learning in the hype of big data analytics. In Pankaj Kumar Sa, Manmath Narayan Sahoo, M. Murugappan, Yulei Wu, and Banshidhar Majhi, editors, *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*, pages 3–25, Singapore, 2018. Springer Singapore.

[4] Varun Ojha, Ajith Abraham, and Vaclav Snasel. Metaheuristic design of feedforward neural networks: A review of two decades of research. *Engineering Applications of Artificial Intelligence*, 60:97–116, 04 2017.

[5] Vina Ayumi, L. M. Rasdi Rere, Mohamad Ivan Fanany, and Aniati Murni Arymurthy. Optimization of convolutional neural network using microcanonical annealing algorithm, 2016.

[6] L.M. Rere, Mohamad Ivan Fanany, and Aniati Arymurthy. Metaheuristic algorithms for convolution neural network. *Computational Intelligence and Neuroscience*, 2016, 05 2016.

[7] Zhonghuan Tian and Simon Fong. *Survey of Meta-Heuristic Algorithms for Deep Learning Training*. 2016.

[8] Bin Wang, Yanan Sun, Bing Xue, and Mengjie Zhang. Evolving deep neural networks by multi-objective particle swarm optimization for image classification. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, page 490–498, New York, NY, USA, 2019. Association for Computing Machinery.

[9] Mohamed Abd El Aziz, Ahmed A. Ewees, and Aboul Ella Hassanien. Multi-objective whale optimization algorithm for content-based image retrieval. *Multimedia Tools and Applications*, 77(19):26135–26172, Oct 2018.