AN IMPROVED GREY WOLF OPTIMIZER WITH LVY WALK AS AN OPTIMIZER IN TRAINING MULTI-LAYER PERCEPTRON WITH DECOUPLED NEURAL INTERFACE

Ahmad Affendi Hashim   Rosni Abdullah

December 5, 2018

Abstract

Grey Wolf Optimizer (GWO) is inspired by how grey wolves (Canis Lupus) searching its prey. The GWO relatively new swarm-based intelligence and the only algorithms that are based on the leadership hierarchy. In GWO, four types of grey wolves such as alpha, beta, delta and omega are employed simulating the leadership hierarchy. Additionally, there are three main steps of hunting, searching for prey, encircling prey and attacking prey are implemented. To improve the GWO search ability, this study proposed Lvy - GWO based on Lvy walk. Five well define benchmark functions were selected in this study. The five benchmark functions were selected based on its features that have many local minima. The results indicate that Lvy -GWO did improve the original GWO based on the error value. Based on Lvy - GWO algorithm. It will be then proposed serving as an optimizer in training multi-layer perceptron (MLP) with Decouple Neural Interface (DNI).

1 INTRODUCTION

Metaheuristics optimizations techniques have become a very popular in the last few decades. The concept of metaheuristics is mostly based on a fairly simple concept and is typically inspired by physical phenomena, animals, behaviors or evolutionary concepts [1].

Many swarm-based optimization algorithms have been developed based on various intelligence creature such as ant, wolf, honey bees, birds, whales etc. have been proven its ability to deal with non-linear, non-convex, discontinuous and discrete optimization problems [2].

GWO is based on swarm-based intelligence and have been proven to have great potential to deal with real world optimization problems [3, 4, 5, 6, 7, 8, 9].

Lvy distribution is the foundation of the Lvy walk (or flight). Lvy walk started gaining a lot of interest around 1999, when [10] published an article in Nature showing it an optimal search strategy for finding sparsely, randomly distributed targets. A stunning variety of organisms, including bacteria, flies, monkeys and sharks, have been described as performing Lvy walks. To perform a Lvy walk, the walker selects a random direction in space, and moves in that direction in a straight line. Depending on how you formulate the walk, any of the duration, length or speed of the walker is drawn from a Lvy distribution.

Training directed neural networks typically requires forward-propagating data through a computation graph, followed by backpropagating error signal, to produce weight updates. All layers, or more generally, modules of the network are therefore locked, in the sense that they must wait for the remainder of the network to execute forwards and propagate error backwards before they can be updated [11]. In DNI, the weights are updated by predicting the value of gradients (which is called synthetic gradients) using only local information. This has allowed the backpropagation process to be eliminated and therefore greatly improve computational cost in terms of computational time and hardware resources.

2 PROPOSED WORK

To meet the objectives of this research, several proposed works has been proposed. First, the proposed GWO and Lvy-GWO will be benchmark and the performance will be compared. Second an improved Lvy GWO will be benchmark and the performance will be compared against backpropagation with gradient descent and DNI. Third, the proposed Lvy-GWO with DNI or Lvy-GWO-DNI will be benchmark and compared against the result gathered in the second objective.

3 PRELIMINARY RESULTS

GWO algorithm is inspired by how Grey wolves hunting its prey. Grey wolves mostly prefer to live in a pack with the group size between 5-12 wolves on average [1, 2]. Social hierarchy is the main feature of this pack. There are four types of wolves which is alpha wolf (α), beta wolf (β), omega wolf (ω) and delta wolves (δ). The alpha wolf is the leader and responsible for all decisions of the pack. The beta wolf is the second category of wolf that responsible helping the alpha wolf in decision making. The third category of wolf is the delta wolf that helps to prevent internal problems of fighting.

By analyzing the social behavior of wolves pack the candidate having best fit is considered as alpha wolf or α solution. Candidates having the second best and third best fit are called as Beta wolf or β solution and delta wolf or δ solution and the remaining solutions are considered as Omega wolves or ω Solutions. The ω Solutions are iteratively improved by following other leading wolves.

The drawback of the GWO algorithm is, all the wolves in the pack will update their position based on the leading wolves of the pack. This has raised the natural question which the wolves in the pack should take guidance from the alpha wolf and not the beta, delta or omega wolves. This is the main drawbacks and could be the main caused why the pack not converge to global optima.

To solve this problem this study proposed an improved random walk called Lvy walk applied to the alpha, beta and delta wolves. An experiment has been carried out to test an improvement gain by implementing Lvy walk. Five benchmark functions have been proposed (refer to Table 1). These five functions were chosen because of its features that contain many local optima. Each of these five functions was run five times and an error for each value was recorded.

GWO and Lvy-GWO were tested using the benchmark functions on table 1 with 10 and 30 dimensions. The results indicate that Lvy-GWO did improve compared to GWO (refer table 2 and 3). However, based on table 2 and table 3, it is observed that both Lvy-GWO and GWO still suffer high-dimensionality problem.

References

[1] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[2] S. Gupta and K. Deep, "A novel Random Walk Grey Wolf Optimizer," *Swarm and Evolutionary Computation*, jan 2018.

[3] A. Madadi and M. M. Motlagh, "Optimal Control of DC motor using Grey Wolf Optimizer Algorithm," *Technical Journal of Engineering and applied*, pp. 373–379, 2014.

[4] N. Muangkote, K. Sunat, and S. Chiewchanwattana, "An improved grey Wolf optimizer for training q-Gaussian Radial Basis Functional-link nets," *2014 International Computer Science and Engineering Conference, ICSEC 2014*, pp. 209–214, 2014.

[5] X. Song, L. Tang, S. Zhao, X. Zhang, L. Li, J. Huang, and W. Cai, "Grey Wolf Optimizer for parameter estimation in surface waves," *Soil Dynamics and Earthquake Engineering*, vol. 75, pp. 147–157, 2015.

[6] A. A. El-Fergany and H. M. Hasanien, "Single and Multi-objective Optimal Power Flow Using Grey Wolf Optimizer and Differential Evolution Algorithms," *Electric Power Components and Systems*, vol. 43, pp. 1548–1559, aug 2015.

[7] S. Mirjalili, "How effective is the Grey Wolf optimizer in training multi-layer perceptrons," *Applied Intelligence*, vol. 43, no. 1, pp. 150–161, 2015.

[8] M. R. Shakarami and I. Faraji Davoudkhani, "Wide-area power system stabilizer design based on Grey Wolf Optimization algorithm considering the time delay," *Electric Power Systems Research*, vol. 133, pp. 149–159, 2016.

[9] D. Guha, P. K. Roy, and S. Banerjee, "Load frequency control of interconnected power system using grey Wolf optimization," *Swarm and Evolutionary Computation*, vol. 27, pp. 97–115, 2016.

[10] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. Da Luz, E. P. Raposo, and H. E. Stanley, "Optimizing the success of random searches," *Nature*, vol. 401, no. 6756, pp. 911–914, 1999.

[11] M. Jaderberg, W. M. Czarnecki, S. Osindero, O. Vinyals, A. Graves, D. Silver, and K. Kavukcuoglu, "Decoupled Neural Interfaces using Synthetic Gradients," vol. 1, aug 2016.

Appendix

Table 1: Mathematical Formulae of the Benchmark Functions.

Name	Mathematical Formulae
Ackley	$f(x_0 \cdots x_n) = -20exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) - exp(\frac{1}{n}\sum_{i=1}^{n} cos(2\pi x_i)) + 20 + e$ $Where -32 \leq x_i \leq 32$ minimum at $f(0, \cdots, 0) = 0$
Rastrigin	$f(x_1 \cdots x_n) = 10n + \sum_{i=1}^{n}(x_i^2 - 10cos(2\pi x_i))$ $Where - 5.12 \leq x_i \leq 5.12$ minimum at $f(0, \cdots, 0) = 0$
Schwefel	$f(x_1 \cdots x_n) = \sum_{i=1}^{n}(-x_i sin(\sqrt{\lvert x_i \rvert})) + \alpha \cdot n$ $Where\ \alpha = 418.982887$ $-512 \leq x_i \leq 512$ minimum at $f(420.968746, 420.968746, \cdots, 420.968746) = 0$
Styblinski & Tang	$f(x) = f(x_1, ..., x_n) = \frac{1}{2}\sum_{i=1}^{n}(x_i^4 - 16x_i^2 + 5x_i)$ $Where - 5 \leq x_i \leq 5$ minimum at $f(x^*) = -39.16599\ at\ x^* = (-2.903534, \cdots, -2.903534)$
Lvy	$f(x, y) = sin^2(3\pi x) + (x-1)^2(1 + sin^2(3\pi y)) + (y-1)^2(1 + sin^2(2\pi y))$ $Where - 10 \leq x_i \leq 10$ minimum at $f(x^*) = 0\ at\ x^* = (1, \cdots, 1)$

Table 2: Average, Standard Deviation, Median, Minimum, and Maximum Error Value Obtained by GWO and Lvy-GWO for 10-Dimentional.

Function	Algorithm	Mean	Std. Dev.	Median	Minimum	Maximum
Ackley	GWO	2.102495	3.168036	0.153160	0.000005	12.575494
	Lvy-GWO	0.004571	0.006149	0.000926	0.000000	0.025631
Rastrigin	GWO	0.333264	0.458114	0.036082	0.000459	1.007058
	Lvy-GWO	0.000000	0.000000	0.000000	0.000000	0.000000
Schwefel	GWO	345.309354	288.641078	351.151407	0.738600	920.968700
	Lvy-GWO	324.548010	273.565424	278.390433	0.272665	795.860180
Styblinski & Tang	GWO	1.707613	2.333746	0.036469	0.001535	5.675073
	Lvy-GWO	1.510281	2.198881	0.106733	0.000579	5.736688
Levy	GWO	0.000138	0.000219	0.000062	0.000000	0.001210
	Lvy-GWO	0.000081	0.000284	0.000000	0.000000	0.001456

Table 3: Average, Standard Deviation, Median, Minimum, and Maximum Error Value Obtained by GWO and Lvy-GWO for 30-Dimentional.

Function	Algorithm	Mean	Std. Dev.	Median	Minimum	Maximum
Ackley	GWO	2.355670	3.603498	0.326560	0.001792	12.607918
	Lvy-GWO	0.028989	0.051345	0.000567	0.000008	0.362579
Rastrigin	GWO	0.419499	0.532169	0.072865	0.000758	1.996776
	Lvy-GWO	0.000000	0.000000	0.000000	0.000000	0.000000
Schwefel	GWO	401.818669	193.808557	417.712496	0.094674	920.968700
	Lvy-GWO	385.254972	252.298695	398.345495	0.694297	920.968700
Styblinski & Tang	GWO	2.275447	2.118585	2.621894	0.001815	5.775527
	Lvy-GWO	2.140668	2.111256	2.439622	0.001811	5.775527
Levy	GWO	0.000165	0.000263	0.000097	0.000000	0.002194
	Lvy-GWO	0.000068	0.000236	0.000004	0.000000	0.002133