Available Online at http://journalijcar.org

International Journal of Current Advanced Research
Vol 5, Issue 7, pp 1074-1077, July 2016

ISSN: 2319-6505

International Journal of Current Advanced Research

ISSN: 2319 - 6475

RESEARCH ARTICLE

AN ANALYSIS OF METAHEURISTIC ALGORITHMS FOR OPTIMIZATION WITH DATA CLUSTERING IN DATA MINING FOR CLASSIFICATION

RashmiAmardeep*[1] and K ThippeSwamy[2]

[1]Sri Siddhartha Academy of Higher Education, Tumkur, India
[2]Visvesvaraya Technological University, PG Regional Center Mysore, India

A R T I C L E I N F O

Article History:

Received 10th April, 2016
Received in revised form 09th May, 2016
Accepted 25th June, 2016
Published online 25th July, 2016

Key words:

Bat Algorithm, Swarm Intelligence, Optimization, Ant Colony Algorithms, Bio-inspired Computing

A B S T R A C T

A metaheuristic provides a sufficiently good solution to an optimization problem. In this paper, we propose few metaheuristic Algorithms. In this paper, we give the pros and cons of PSO, ACO, and BA algorithm. We show the superiority of the new metaheuristic bat algorithm (BA) over other standard algorithms such as ACO and Particle Swarm Optimization. For comparison purpose, we have studied and analyzed few Swarm Intelligence (SI) algorithms on different datasets. The experiments are conducted on attributes which are categorical and continuous for the Fi classi cation. attribute are continuous. The performance of every classifier was evaluated by calculating the weighted Arithmetic mean, Normalized Absolute error, Standard Deviation and accuracy. The experimental results show that the BAT classifier performs better than all the compared algorithms.

INTRODUCTION

A metaheuristic is a high-level problem- independent algorithmic framework that provides a set of guidelines to develop heuristic optimization algorithms. The vast majority of heuristic and metaheuristic algorithms have been derived from the behavior of biological systems and/or physical systems in nature. Some of the metaheuristics include Swarm intelligence [7, 8], tabu search, simulated annealing [9]. Swarm intelligence has been successfully used to solve optimization problems in the engineering, the financial, and the management fields. For example, particle swarm optimization (PSO) techniques have successfully been used to construct the portfolios of stock. [1] Ant colony optimization (ACO) techniques have successfully been used to solve the traveling salesman problem (TSP) [3] and the routing problem of networks [5]. These algorithms have certain advantages and disadvantages.

BA is a relatively new population-based metaheuristic approach based on hunting behavior of microbats. The capability of echolocation of microbats is fascinating as these bats can _nd (fi) their prey and discriminate di erent types of insects even in complete darkness. We will _rst (fi) formulate the bat algorithm by idealizing the echolocation behavior of bats. We then describe how it works and make a comparison with other existing algorithms.

Particle Swarm optimization (PSO)

Dr. Eberhart and Dr. Kennedy in 1995 developed Particle swarm optimization (PSO) is a population-based stochastic optimization technique. Instead of using evolutionary operators to manipulate the individuals, like in other evolutionary computational algorithms, each individual in PSO [3, 7] flies in the search space with a velocity which is dynamically adjusted according to its own flying experience and its companions' flying experience.

The ith particle is represented as Xi = (xil, xi2, ..., xid). The best previous position (the position giving the best fitness value) of the ith particle is recorded and represented as Pi = (p il, p i2, ….., p id). The index of the best particle among all the particles in the population is represented by the symbol Gbest representing global best. The rate of the position change (velocity) for particle i is represented as Vi to the following equation: (vil, vi2... vid). Concept of modification of a searching point by PSO is shown in Figure (1).The particles are manipulated according to the following equations:

$$Vi^{k+1}= w*V_i^k +c_1 *rand1*(Pbest_i *x_i k) +c_2 * rand2 *(G_{best} - x_i^k) \text{ ---- } (1)$$

$$X_i^{k+1} = x_i^k +v_i^k \text{------ } (2)$$

Where,

k - Dimension of the problem space
rand1, rand2 - random values in the range of (0, 1)
c1, c2 - acceleration coefficients constants
w - Inertia weight factor

The pseudo code of the procedure is as follows:

For each particle
    Initialize particle
END

Do
For each particle

Calculate fitness value

If the fitness value is better than the best fitness value (pBest) in history set current value as the new pBest

End

Choose the particle with the best fitness value of all the particles as the g Best For each particle

Calculate particle velocity according equation (1)

Update particle position according equation (2)

End

While maximum iterations or minimum error criteria is not attained.

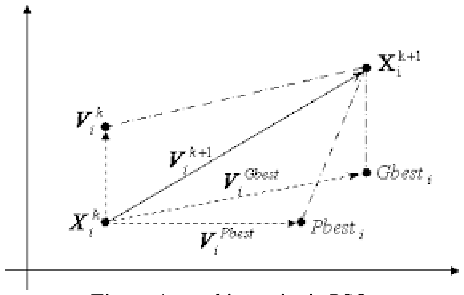

Figure 1 searching point in PSO

Advantages

- Simple, easy and derivative free algorithm.
- Very few parameters to adjust and efficient global search ability
- Efficient to handle complex non-linear optimization problems.
- They have internal memory and each particle represents individual solution.
- Best information sharing mechanism.

Disadvantages

- For large search space, premature convergence to local optima
- Weak local search

Ant Colony Optimization (ACO)

Ant colony optimization algorithm is metaheuristic Soft computing technique for solving hard discrete optimization problems [11] [12]. Initially proposed by Marco Dorigo in 1992, based on the behavior of ants seeking a path between their colony and a source of food.

Procedure ACOMeta Heuristic

While (not_termination)
GenerateSolutions ()
DaemonActions ()
PheromoneUpdate ()
Endwhile
EndProcedure

The probability that the k-th ant will choose the city j as its next travel point is defined by a probability function. This function applied for ant k currently at city i during iteration t is of the form

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{k \in allowed_k}[\tau_{ik}(t)]^{\alpha}[\eta_{ik}]^{\beta}} & \text{if } j \in allowed_k \\ 0 & \text{otherwise} \end{cases} \quad --- (3)$$

The pheromone is updated as the ant moves from one city to another and represents the learned desirability of choosing city j when in city i. This update is performed as follows:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho\Delta\tau_{ij}^{best}(t)$$

To write our different scripts, we focused on the following parameters: the exponents α and β used in the probabilistic decision rule (see equation 3) and the evaporation parameter ρ. Alpha=1 Beta=2 Evap=0.5 ant colony size 50

Advantages

- Positive Feedback accounts for rapid discovery of good solutions
- Distributed computation avoids premature convergence
- The greedy heuristic helps find an acceptable solution in the early solution in the early stages of the search process.
- The collective interaction of a Population of agents.

Disadvantages

- Slower convergence than other Heuristics
- No centralized processor to guide the AS towards good solutions

Bat Algorithm

Xin-SheYang proposed BA [15, 16] is a metaheuristic optimization algorithm in 2010. This bat algorithm is based on the echolocation behavior of microbats with varying pulse rates of emission and loudness [17, 18].

In the BA, an artificial bat has a position vector, velocity vector, and frequency vector which are updated during the course of iterations as (4), (5), and (6):

Vi(t+1)=Vi(t)+(Xi(t) Gbest)Fi --(4)
Xi (t+1) =Xi (t) +VI (t+1) -- (5)

Where Gbest is the best solution attained so far and Fi indicates the frequency of i- th bat which is updated in each course of iteration as follows:

Fi=Fmin+ (Fmax Fmin) β – (6)

Where β is a random number of a uniformdistribution in [0, 1]. It is clear from the Eqs. (6) and (6) that different frequencies encourage artificial bats to have diverse propensity to the best solution.

These equations could guarantee the exploitability of the BA. However, a random walk procedure is also has been used to perform the exploitation as follows:

Xnew=Xold+εA t

In this formula, ε is a random number in [1,1], and A is the loudness of emitted sound that bats use to perform an exploration instead of exploitation as it is increased. According to these concepts, it can be stated that BA is a balanced combination of PSO and intensive local search. The balancing between these techniques is controlled by the loudness (A) and pulse emission rate (r). These two elements are updated as follows:

$A_i^{t+1}=\alpha A_i^t$ --- (7)
$r_i^{t+1}=r_i^0[1 \ \exp(\ \gamma t)]$ ----- (8)

Where α and γ are constants. Eventually, A i will equal zero, while the final value of r i is r (0). Note that both loudness and rate are updated when the new solutions are improved to guarantee that the bats are moving toward the best solutions.

RESULT ANALYSIS

The proposed research work is designated to perform comparative analysis on well- known datasets in order to test the efficiency of algorithm, PSO, ACO and Bat algorithm. The datasets, which we are intended to use are selected from the real datasets of UCI and are easily available at link http://archive.ics.uci.edu/ml/datasets. html these datasets are namely Breast Cancer Wisconsin, Wine, and Zoo. Each & every considered data sets are having a different number of dimensions/attributes. The detailed characteristics of these data sets [10], [11] are well illustrated in Table I as given below.

Table I Characteristic of Dataset

	Dimension Type	Total no. of instances	No. of Dimension s
Breast Cancer Wisconsin	Integer	683	9
Wine	Integer, Real	150	13
Zoo	Integer	101	17

Each dataset is partitioned into ten data subsets. In each run a different partition is used as a testing set and the remaining 9 are grouped together to build training set. The training set is used to train the model for good learning capability, in which the generalization capability of the proposed classifier is evaluated by the testing set. The performance is measured on the basis of mean, standard deviation and normalized absolute error (NAE) and accuracy. Arithmetic Mean is calculated to measure the central tendency of data. Figure 1 shows the calculated weighted arithmetic mean values by using three optimization algorithms Bat, Ant Colony Optimization and Particle Swarm Optimization Algorithm respectively.

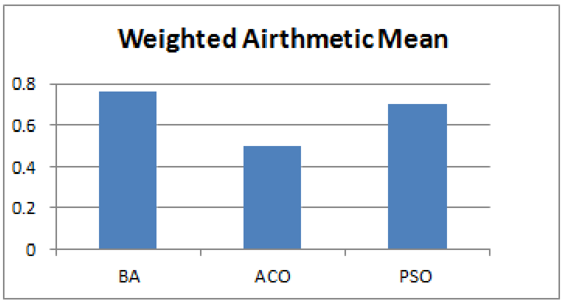

Figure 1 Arithmetic mean

The standard deviation (σ) shows how much variation or dispersion from the average exists. Standard Deviation is calculated to measure the dispersion of data. Figure 3 shows the calculated standard deviation values by using three optimization algorithms Bat, Ant Colony Optimization and Particle Swarm Optimization Algorithm respectively.

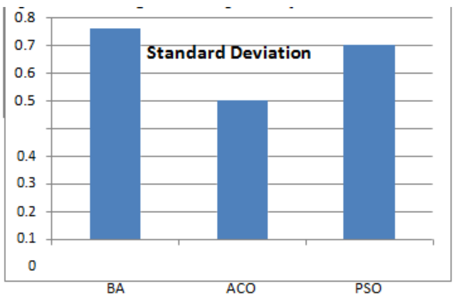

Figure 2 Standard Deviation

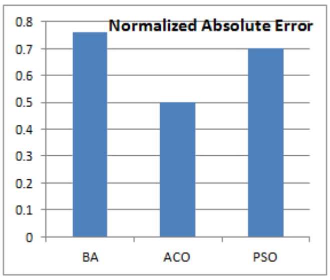

Figure 3 Normalized Absolute Error

The classification accuracy is defined as follow:

$$\text{Accuracy} = \frac{\text{Number of correctly classified objects}}{\text{Number of objects in datasets}} * 100\ \%$$

Normalization is the process of isolating the statistical error in repeated measured data. These ratios only make sense for ratio measurements in terms of levels of measurement. Figure 3 shows the calculated normalized absolute error values by using three optimization algorithms Bat Algorithm, Ant Colony Optimization and Particle Swarm Optimization Algorithm respectively. The accuracy in percentage of the algorithm is in Table II

Table II Accuracy of algorithm

Data Sets	PSO	ACO	BA
WBC	95.150	96.13	96.25
Zoo	95.480	96.06	96.78
Wine	93.78	95.45	95.12

CONCLUSION AND FUTURE SCOPE

We have analyzed the different aspects of the applicability of several SI. Calculating the weighted Arithmetic mean, Normalized Absolute error, Standard Deviation and accuracy are important aspects to be considered for comparison purpose. The SI will be tested on different kinds of datasets having multiple dimensions. From the results, it seems that the Bat algorithm has a good performance of predictive accuracy. Future work is to investigate the use of Bat algorithm along with back-propagation neural network (BPNN) algorithm to solve the local minima problem and also to enhance the convergence rate

References

1. J.-F. Chang and S.-W. Hsu: The Construction of Stock's Portfolios by Using Particle Swarm Optimization. Proceedings of the 2nd International Conference on Innovative Computing, Information and Control, (2007).
2. S.-C. Chu and P.-W. Tsai: Computational Intelligence Based on the Behavior of Cats, *International Journal of Innovative Computing, Information and Control* 3(1), 163 (2006).
3. M. Dorigo and L. M. Gambardella: Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem, IEEE Transactions on Evolutionary Computation 1(1), 53 (1997).
4. C. P. Pinto, A. Nägele, M. Dejori, T. A. Runkler, and J. M. C. Sousa: Using a Local Discovery Ant Algorithm for Bayesian Network Structure Learning, IEEE Transactions on Evolutionary Computation 13(4), 767 (2009).

5. Kennedy, J and Eberhart, R.: Particle swarm optimization, Proc. IEEE Int. Conf. Neural Networks. Perth, Australia, 1942- 1945 (1995).
6. Kennedy, J and Eberhart, R., Swarm Intelligence. Academic Press, (2001). [9] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by simulated annealing. Science, 220, 671-680 (1983).
7. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by simulated annealing. Science, 220, 671-680 (1983).
8. S.Rana, S. Jasola and R. Kumar. (2012, June). A boundary restricted adaptive particle swarm optimization for data clustering. *International Journal of Machine Learning & Cyber. Springer.* [Online]. pp. 391-400. Available: link. springer.com/ content/ pdf/10.1007/s13 042-012-0103-y.pdf
9. Mariam El-Tarabily, Rehab Abdel- Kader, Mahmoud Marie and Gamal Abdel-Azeem. (2013, Aug.). A PSO- Based Subtractive Data Clustering Algorithm. *International Journal of Research in Computer Science*. [Online].3(2).pp. 1-9. Available: http://www.ijorcs.org/manuscript/id/60/doi:10. 7815/ ijorcs.32.2013.060/ma riam-el- tarabily/a-pso-based-subtractive-data- clustering-algorithm
10. D. Martens, M. De Backer, R. Haesen, J. Vanthienen, M. Snoeck, B. Baesens,Classification with Ant Colony Optimization, IEEE Transactions on Evolutionary Computation, volume 11, number 5, pages 651—665, 2007.
11. M, den Bseten, T. Stützle and M. Dorigo, "Ant colony optimization for the total weighted tardiness problem," Proceedings of PPSN-VI, Sixth International Conference on Parallel Problem Solving from Nature, vol. 1917 of Lecture Notes in Computer Science, pp.611-620, 2000.
12. M. Dorigo T. Stützle, "The AntColony Optimization Metaheuristic: Algorithms, Applications, and Advances", Handbook of Metaheuristics, 2002
13. X. S. Yang. Bat algorithm for multi- objective optimisation. *International Journal of BioInspired Computation,* 3(5):267-274, 2011.
14. X. S. Yang. A new metaheuristic bat- inspired algorithm. Nature Inspired Cooperative Strategies for Optimisation (NICO 2010), pages 65-74, 2011.
15. A.H.Gandomi, X. S. Yang, A. H. Alavi and S. Talatahari. Bat algorithm for constrained optimisation tasks. Neural Computing and Applications, pages 1-17, 2012.
16. P. W. Tsai, J. S. Pan, B. Y. Liao, M. J. Tsai and V. Istanda. Bat inspired algorithm for solving numerical optimisation problems. Applied Mechanics and Materials, 148:134137, 2012.

****