ISSN 2229-5518

Comparison on Swarm Algorithms for Feature Selections/Reductions

Mohammad Aizat bin Basir, Faudziah binti Ahmad

Abstract— The swarm algorithms that are stimulated by the principles of natural biological evolution and distributed collective behavior of social colonies have shown excellence in dealing with complex optimization problems and are becoming more popular nowadays. This paper surveys six (6) biologically inspired popular swarm algorithms, namely Particle Swarm Optimization (PSO), Artificial Ant Colony Optimization Algorithms (ACO), Artificial Fish Swarm Algorithms (AFSA), Artificial Bees Colony Algorithms (ABC), Firefly Algorithms (FA) and Bat Algorithms (BA) and its application in feature selections/reductions. The process of reviewing articles is by looking into how the algorithms adapted to the problem in feature selections/reductions specifically in improving the classification accuracy, minimizing the numbers of attributes without compromising the quality of the results and evaluating performance in term of processing time. Results shows that all algorithms reviewed based on previous literature have a promising capability that can be applied in feature selections/reductions problem. The significance of this reviewed is to present the comparison and various alternative of swarm algorithms to be applied in feature selections/reductions problem.

Index Terms— Feature Selection/Reduction, Particle Swarm Optimization (PSO), Artificial Ant Colony Optimization Algorithms (ACO), Artificial Fish Swarm Algorithms (AFSA), Artificial Bees Colony Algorithms (ABC), Firefly Algorithms (FA), Bat Algorithms (BA).

1 INTRODUCTION

Swarm-based algorithms are inspired by the behaviour of some social living beings, such as ants, bats, bees, fireflies and fishes. The important features of swarm-based system are self-organization and decentralized control lead to an emergent behaviour. Emergent behaviour is a property that emerges through local interactions among system components and it is impossible to be achieved by any of the components of the system acting alone. In the previous years, the two mainstreams of the Swarm Intelligence area were ant colony optimization [1] and particle swarm optimization (PSO)[2]. In the recent years, new swarm intelligence algorithms have appeared, inspired by fish schools [3], as well as different aspects of the behaviour of bees [4]–[7], bacteria [8], glow-worms [9], fireflies [10], cockroaches [11], bats [12], and cuckoo birds [13]. Despite the swarm inspiration common to these approaches, they have their own particular way to exploit and explore the search space of the problem.

Firstly, the methodology of constructing the paper was introduced. Then, the result of the paper review been presented including subsections: 1) Evolution of swarm algorithms 2) A brief introduction of swarm algorithms together with advantages and disadvantages of six (6) popular swarm algorithms namely Particle Swarm Optimization (PSO), Artificial Ant Colony Optimization Algorithms (ACO), Artificial Fish Swarm Algorithms (AFSA), Artificial Bees Colony Algorithms (ABC), Firefly Algorithms (FA) and Bat Algorithms (BA) 3) A brief discussion of swarm algorithms application in feature selections/reductions 4) Lastly the discussion and conclusions briefly discussed.

- *Mohammad Aizat bin Basir is currently pursuing doctoral degree program in computer science in Universiti Utara Malaysia, Malaysia. E-mail: uk6565@yahoo.com*
- *Faudziah binti Ahmad is currently Assoc. Prof. in computer science in Universiti Utara Malaysia, Malaysia, E-mail: fudz@uum.edu.my*

2 METHODOLOGY

This is the process of reviewing in this article:

i. Articles from early 90's to current year 2014 were randomly selected based on relevancy to the article. That is, this article look for application of swarm algorithms in feature selections/reductions, particularly how the algorithms adapted to the problem in feature selections/reductions specifically in improving the classification accuracy, minimizing the numbers of attributes without compromising the quality of the results and evaluating performance in term of processing time.

ii. The sources were from Google Scholar search engine and various research online databases.

3 RESULTS

The following sections present information in terms of inventors, purpose, advantages/ disadvantages, and related works on six (6) swarm algorithms that have been adapted to address feature selections/reductions problems. These algorithms include PSO, ACO, AFSA, ABC, FA and BA. The information's gathered are presented according to the evolution of these algorithms in addressing feature selection problems.

3.1 Evolution of swarm algorithms

PSO is a one of the earliest swarm algorithm that been introduced in year 1995 for optimizes a problem by iteratively trying to improve a candidate solution with regard to a given

ISSN 2229-5518

measure of quality [2]. In general, PSO is inspired by social behaviour patterns of organisms that live and interact within large groups. In particular, it incorporates swarming behaviours captured in flocks of birds, schools of fish, or swarms of bees, and even human social behaviour. A basic PSO algorithm works by having a population (called a swarm) of candidate solutions (called particles). Particles are moved around in the search-space according to a few simple formulas. The particles movement is guided by their own best known position in the search-space as well as the entire swarm's best known position. The process is repeated until improved positions are being discovered and satisfied. PSO has demonstrated good performance in various applications [14]. It is important to highlight PSO main strength which is more faster convergence compares with many other global optimization algorithms [15], [16]. In addition, PSO also a meta-heuristic algorithms which makes few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. However, meta-heuristics algorithms such as PSO do not guarantee an optimal solution is ever found.

ACO is another popular algorithms that been applied in various fields of optimization after the PSO been introduced. It's been in an optimization algorithms modelled on the actions of an ant colony [17]. ACO is a probabilistic technique useful in problems that deal with finding better paths through graphs based on behaviour of ants seeking a path between their colony and source of food. Artificial ant (simulation agent) play important role to locate optimal solutions by moving through a parameter space representing all possible solutions. Natural ants lay down pheromones directing each other to resources while exploring their environment. The simulated 'ants' similarly record their positions and the quality of their solutions, so that in later simulation iterations more ants locate better solutions. ACO has known for some advantages which are:

i. Inherent parallelism
ii. Positive Feedback accounts for rapid discovery of good solutions
iii. Efficient for Traveling Salesman Problem and similar problems
iv. Can be used in dynamic applications (adapts to changes such as new distances)
v. Distributed computation avoids premature convergence.

Because ACO is a heuristic algorithm, ACO has disadvantages of defects of searching local optimization and slow convergence speed [17]. Another disadvantages that been stated in the literature are:

i. Theoretical analysis is difficult
ii. Sequences of random decisions (not independent)
iii. Probability distribution changes by iteration
iv. Research is experimental rather than theoretical
v. Time to convergence uncertain.

Three years later, an artificial fish swarm algorithm (AFSA) was proposed by Li (2003) in his doctoral thesis based on the basic idea to imitate the fish behaviours such as praying, swarming, and following with local search of fish individual for reaching the global optimum [18]. In water areas, a fish can always find food at a place where there are plenty of food by following other fishes, hence generally the more the food, the more the fish. Following this rule, artificial fish school algorithm (AFSA) builds some artificial fish (AF), which search an optimal solution in solution space (the environment in which AF live) by imitating fish swarm behaviour. Three basic behaviours of AF are defined as follows [3]:

i. Prey: The fish perceives the concentration of food in water to determine the movement by vision or sense and then chooses the tendency.
ii. Swarm: The fish will assemble in groups naturally during the moving process, which is a kind of living habits in order to guarantee the existence of the colony and avoid dangers.
iii. Follow: In the moving process of the fish swarm, when a single fish or several fishes find food, the neighbourhood partners will trail and reach the food quickly.

AFSA has an advantages of possess similar attractive features of genetic algorithm (GA) such as independence from gradient information of the objective function, the ability to solve complex nonlinear high dimensional problems. Furthermore, it can achieve faster convergence speed and require few parameters to be adjusted. Whereas the AFSA does not possess the crossover and mutation processes used in GA, so it could be performed more easily. However, AFSA has a few disadvantages which are:

i. High time complexity
ii. Lack of balance between global and local search.
iii. Lack of benefiting from the experiences of group members for the next movements.

Later in 2005 an artificial bee colony (ABC) algorithm was proposed by Karaboga (2005) for solving multidimensional and multimodal optimization problems [5]. The bees' aim is to discover places of food sources (regions in the search space) with high amount of nectar (good fitness). There are three types of bees: the scout bees that randomly fly in the search space without guidance, the employed bees that exploit the neighbourhood of their locations, selecting a random solution to be perturbed, and the onlooker bees that use the population fitness to select probabilistically a guiding solution to exploit its neighbourhood. If the nectar amount of a new source is higher than the previous one in their memory, they update the new position and forget the previous one (greedy selection). If a solution is not improved by a predetermined number of trials, then the food source is abandoned by the corresponding employed bee and it becomes a scout bee. ABC has been successfully applied in various field of optimization and projected some advantages:

i. Simplicity, flexibility and robustness [11], [19]
ii. Use of fewer control parameters compared too many other search techniques [20].
iii. Ease of hybridization with other optimization algorithms [19].

iv. Ability to handle the objective cost with stochastic nature [21].
v. Ease of implementation with basic mathematical and logical operations.

Despite with excellent advantages, ABC algorithms also comes with the disadvantages which convergence performance of ABC for local minimum is slow.
4 years later after ABC algorithms been invented, Firefly algorithm (FA) was invented by Yang (2008) which is inspired by biochemical and social aspects of real fireflies [22]. Real fireflies produce a short and rhythmic flash that helps them in attracting (communicating) their mating partners and also serves as a protective warning mechanism. FA formulates this flashing behaviour with the objective function of the problem to be optimized. FA has two major advantages over other algorithms: automatically subdivision and the ability of dealing with multimodality. First, FA is based on attraction and attractiveness decreases with distance. This leads to the fact that the whole population can automatically subdivide into subgroups, and each group can swarm around each mode or local optimum. Among all these modes, the best global solution can be found. Second, this subdivision allows the fireflies to be able to find all optima simultaneously if the population size is sufficiently higher than the number of modes. Mathematically, 1/√ controls the average distance of a group of fireflies that can be seen by adjacent groups. Therefore, a whole population can subdivide into subgroups with a given, average distance. In the extreme case when = 0, the whole population will not subdivide. This automatic subdivision ability makes it particularly suitable for highly nonlinear, multimodal optimisation problems. In addition, the parameters in FA can be tuned to control the randomness as iterations proceed, so that convergence can also be sped up by tuning these parameters. These above advantages make it flexible to deal with continuous problems, clustering and classifications, and combinatorial optimisation as well. Despite with excellent advantages, FA has shown a few drawbacks which are [23]:

i. Getting trapped into several local optima.
ii. FA performs local search as well and sometimes is unable to completely get rid of them.
iii. FA parameters are set fixed and they do not change with the time.
iv. In addition FA does not memorize or remember any history of better situation for each firefly and this causes them to move regardless of its previous better situation, and they may end up missing their situations.

In 2010 after one year's FA was proposed, bat algorithm (BA) was first presented by Yang (2010). The basic idea behind the BA is that a population of bats (possible solutions) uses echolocation to sense distance and fly randomly through a search space updating their positions and velocities. The bats' flight aims at finding the food / prey (best solutions) [24]. A loudness decay factor acts in a similar role as the cooling schedule in the traditional simulated annealing optimization method, and a pulse increase factor regulates the pulse frequency. As the loudness usually decreases once a bat has found its prey/solution (in order to not to lose the prey), the rate of pulse emission increases in order to raise the attack accuracy. BA has advantages which claimed to provide very quick convergence at a very initial stage by switching from exploration to exploitation. However, BA also comes with drawback which if BA switches to exploitation stage too quickly, it may lead to stagnation after some initial stage.

3.1.1 Summary

In summary, we were discussed the evolution of swarm algorithms starting from earliest algorithm which is PSO followed by ACO, AFSA, ABC, FA and the recent algorithm which is BA. We also highlighted the basic concept of each algorithms been inspired by the biological behavior of the the swarms colony. The strenghs and drawbacks for each algorithms has been pointed out based on previous literature. It is important to mention that each new algorithm evolved aim to provide alternative and improvement from previous algorithms.

3.2. Applications of swarm algorithm for features selections/reductions

In recent years, many reduction methods have been proposed. This section discusses six (6) swarm algorithms that have been applied to address features selection/reductions problems.

3.2.1 PSO in features selections/reductions

Yue et al. (2007)introduced an approach for finding appropriate features based on rough set using PSO [25]. The proposed method discovered the best feature combinations in an efficient way to observe the change of positive region as the particles proceed through the search space. The performance of the method been evaluated with Genetic Algorithm (GA). Result shown that in term of performance, PSO and GA almost got the same number of reducts but PSO used less iterations than GA. They concluded that PSO required shorter time to obtain better results than GA, especially for large scale problems, although its stability needs to be improved in further research.

Ye, Chen, and Liao (2007) have presented a new algorithm for minimum attribute reduction based on Binary Particle Swarm Optimization (BPSO) with Vaccination [26]. Their research started with transformed the problem of minimum attribute reduction into an unconstrained binary optimization problem. Then they defined the suitable fitness function and the equivalence of optimality between the original problems to prove the transformation been done. In the next step, they approached to solve the transformed problem using an improved BPSO algorithm combined with some vaccination mechanism. Experimental results on a number of data sets obtained from the UCI machine learning repository show that the proposed algorithm has a higher possibility of finding a minimum reduction and remarkably outperforms some existing algorithms specifically designed for minimum attribute reduction in both quality of solution and computational complexity.

Selvaraj and Janakiraman (2013) proposed an improved feature selection based on BPSO for liver disease diagnosis [27]. In this research, they applied BPSO to get the best reduced feature set from auto covariance features that extracted from the segmented lesion. They used BPSO as a feature selector and Probabilistic Neural Network (PNN) as a classifier and claimed that very effectively integrated. The selected best features from BPSO are fed to PNN classifier to classify the liver diseases. Results shown indicate that the time complexity to train the neural network is minimized after reducing the feature set using BPSO. Experiment also demonstrated that by varying the number of particles, iterations and transfer function, the best transfer function can be determined.

In this part, we discussed the work done using PSO as an algorithm to solve feature selections/reductions problem. PSO demonstrated promising results when finding a good feature selection. The advantage highlighted using PSO is compared with the other algorithms, the method is very simple, easily completed and it needs fewer parameters, which made it fully developed. Despite with advantages mentioned, PSO also has a drawbacks which are [28]:

i. If all the velocity becomes equal to V_{max} the particle will continue to conduct searches within a hypercube and will probably remain in the optima but will not converge in the local area.
ii. Achieve optimality convergence strongly influenced by the inertia weight.
iii. When the algorithm converges, the fixed values of the parameters might cause the unnecessary fluctuation of particles.
iv. Higher throughput: More sophisticated finite element formulations, higher accuracy (mesh densities).

3.2.2 ACO in features selections/reductions

New heuristic approach for solving the minimal attribute reduction problem (MARP) based on the ant colony optimization (ACO) meta-heuristic has been proposed [29]. They developed a new algorithm R-ACO for solving the MARP and the simulation results claimed that their approach can find more minimal attribute reductions more efficiently in most cases. Basically their research improved previous works [30], [31]. The improvement involved reducing the time cost operation by proposed a new model R-Graph to solve the MARP with ACO. With this approach, they solved the problem with increased more reductions especially towards achieving minimal reductions.

An attribute reduction mechanism that based on Ant Colony Optimization algorithm and rough set theory called (ACOFS) has been proposed [32]. They assessed the effectiveness of their approach through the number of attributes in each reduct and the run time from program start to termination. The performance of the proposed approach was evaluated and compared with three latest algorithms (IDSRSFS, RSFSACO, ARWSO) [33]–[35]. Based on experiment results, they claimed their algorithm performed equally better with three algorithms when handling dataset than has less than 16 attributes. They also found that one the previous algorithm (IDSRSFS) failed to find the minimal reducts as the other algorithms when dataset is more than 16 attributes. They concluded the problem is due to IDSRSFS's drawback not having heuristic information to search through the feature space for optimal solutions and premature convergence to a local optimum in the space. ACOFS algorithm also found to be the fastest algorithm in finding the final results between the reported algorithms, despite the fact that ARWSO and RSFSACO initiate the solution based on the core attributes and they proved the efficiency of ACOFS. As a future work, they suggest to construct ACOFS with changing the solution construction mechanism by employing a heuristic algorithm and considering the core features.

In this section, a numbers of research in feature selections/reductions using ACO have been discussed. ACO found to be a predominantly a useful tool and modern algorithm that has been used in many studies for selecting relevant features. The example of other researches using ACO in feature selections/reductions can be referred [36]–[38]. However, one of the crucial challenges in feature selection is to find a solution in the full search space on the basis of activities of multi-agent systems that use a global search ability utilizing local search appropriately. Because of that researchers simulated ACO [39] as an attempt to achieve global search for finding high-quality solutions within a reasonable period of time. However, there are few common problems of ACO in feature selections/reductions. One of the disadvantages of ACO is most of the ACO-based FS algorithms do not consider the random and probabilistic behaviour of ants during subset constructions. Consequently, the solutions found in these algorithms might be incomplete in nature. There is a need in the future of more hybridization of techniques between ACO and other algorithms to achieve optimal reductions results rather than rely on the ACO independently.

3.2.3 ABC in features selections/reductions

A novel hybridization of rough set theory with Bee Colony Optimization (BCO) has been proposed by [40]. The method proposed did not use any random parameter and claimed their algorithm provide consistent performance. One year after that, an improved Rough Set-based Attribute Reduction (RSAR) namely Independent RSAR hybrid with Artificial Bee Colony (ABC) algorithm has been introduced [41]. They grouped the instances based on decision attributes. Then, they applied Quick Reduct Algorithm [42] to find the reduced feature set for each class. To this set of reducts, they utilized ABC algorithm to select a random number of attributes from each set, based on the RSAR model, to find the final subset of attributes. An experiment was carried out on five different datasets from the UCI machine learning repository. The performance of the reduct is analysed with Genetic k-Nearest Neighbour (GKNN) classifier and compared with six different algorithms (general RSAR, Entity based Reduct (EBR), Genetic RSAR, Ant RSAR, Particle Swarm Optimization based RSAR (PSORSAR) and with their previous work (BeeRSAR). They claimed the proposed method can find very minimal reduct than the other existing methods.

Shunmugapriya and Kanmani (2012) proposed a novel feature selection method in which ABC is used to generate the feature subsets and a classifier is used to evaluate the feature subsets generated by ABC. In this method, each employed bee is allocated a feature (food source) and the onlooker tries to make all possible combinations with other features to configure the feature subset [43]. The proposed algorithm has shown competitive performance compared to ACO [44] based feature selection.

Schiezaro and Pedrini (2013) presented a new feature selection method based on ABC algorithm. After evaluated through 10 data sets from different knowledge fields, the result shows that a reduced number of features can achieve classification accuracy superior to that using the full set of features [45]. They claimed the proposed method presented better results for the majority of the tested data sets compared to other algorithms. They also suggested for future work to investigate alternative mechanisms to explore neighbourhood of food sources, parallelize the exploration of employed bees in relation to the food sources, and create a filter approach combining ABC algorithm, entropy, and mutual information.

In this section, we highlights some of researches been done which applying ABC for feature selections/reductions. More research on ABC been applied in various field can be viewed in the paper presented [46]. ABC remains a promising and interesting algorithm, which would continue to be extensively used by researchers across diverse fields. Its potential advantage of being easily hybridized with different meta-heuristic algorithms and components makes it robustly viable for continued utilization for more exploration. Although ABC has great potential, it was clear to the many scientists that some modifications to the original structure are still necessary in order to significantly improve its performance. Combination with other algorithms hopefully will show promising enhancement in the future for feature selection/reductions.

3.2.4 AFSA in features selections/reductions

Zhang et al. (2006) has presented the use of AFSA as a new tool which sets up a neural network (NN), adjusts its parameters, and performs feature reduction, all simultaneously. They combined the feature selection and NN architecture problem into an optimization procedure and employs AFSA to resolve it [47]. Also, they performed the feature selection and evolving NN architecture at the same time by AFSA, which is not based on a fixed network. Results showed that their proposed method were able to optimize network architecture to be kept simple, reduce compute and enhance generalization ability of the resulting classifier.

Wang and Dai (2013) has promoted the artificial fish swarm algorithm used to search the optimal feature subset and the chaotic, feedback mechanisms are introduced to improve the artificial fish swarm algorithm, the excessive intrusion feature rough sets produced in the classification process are simplified to guarantee the simplicity of characteristics and the estimation model for residuals gray level to predicate the early simplified invasion [48]. Their experiment results illustrated that the improved artificial fish swarm algorithm can obtain more optimal feature subsets and increase the abnormal detection accurate rates and speeds which can be widely applied in network security.

In this part, we presented the application of ASFA in feature selections/reductions. It can be seen that a number of different methods have been proposed to approach the optimal solution to feature selection using AFSA including neural network and various stochastic algorithms. AFSA is one of the most appropriate methods for swarm intelligence optimization, capable of solving the problems by inspiration from the movement of fishes and obtains more optimized results compared with other swarm intelligence algorithms. However, AFSA has some disadvantages like falling in local optimum points, advanced convergence and time consuming [49]. There were an attempt to hybrid the AFSA algorithm with other methods such as Particle Swarm Algorithm (PSO) [50], Shuffle Frog Leaping Algorithm (SFL) [51], Cellular Learning Automata (CLA) [52] and continuing to grow with hybrid efforts in order to find the optimal solution for problem selected and overcome the drawbacks in AFSA.

3.2.5 BA in features selections/reductions

BA can deal with highly nonlinear problem efficiently and can find the optimal solutions accurately [12], [24], [53]. Case studies include pressure vessel design, car side design, spring and beam design, truss systems, tower and tall building design and others. It is been solved numerical optimization problems using bat algorithm [54]. In addition, research done to optimized the brushless DC wheel motors using bat algorithm with superior results [55].

Nakamura et al. (2012) has proposed propose a new nature-inspired feature selection technique based on the bats behaviour. The technique implemented wrapper approach combines the power of exploration of the bats together with the speed of the Optimum-Path Forest classifier to find the set of features that maximizes the accuracy in a validating set [53]. Experiment employed five public datasets to accomplish this task, in which BA has been compared against Binary PSO, Binary FFA and Binary GSA. They claimed the proposed algorithm outperformed the compared techniques in 3 out of 5 datasets, being the second best in the remaining two datasets.

In this section, we highlight some of BA application in feature selections/reductions. BA efficiently can be applied in feature selections based on work reviewed. Research work has shown that BA can deal with problem of the high dimensionality and finding the most informative features in a search space. The reason is BA has a capability of automatically zooming into a region where promising solutions have been found which accompanied by the automatic switch from explorative moves to local intensive exploitation. As a result, BA has a quick convergence rate, at least at early stages of the iterations compared with other algorithms. However, there are still some advantages found in BA. The drawbacks of BA including the performance of BA are largely dependent on the parameters of the algorithm, sometime inefficiently parameter controls and low convergence speed. In conclusion, BA still has more opportunities to improve especially on how to vary or control parameter tuning during iterative search process and improve performance of BA related to convergence speed.

3.2.6 FA in features selections/reductions.

FA for feature selection has been proposed and produced consistent and better performance in terms of time and optimality than other algorithms [56]. They claimed that RSAR approach with nature inspired algorithm to improve the performance such as GenRSAR [57], AntRSAR [30], PSO-RSAR [25] and BeeRSAR [40] able to increase the degree of optimality but are not consistent as it varies with the parameter values which are application dependent. They also found that BeeRSAR algorithm consumes more time to find the reduct even though not require any random parameter [40]. In way to improve the disadvantage found in the previous algorithms, they proposed novel approach for feature selection based on nature inspired "Firefly" algorithm. The proposed algorithm (FA_RSAR) was an effort that combines FA together with Rough Set Theory (RST) to ensure the success in less time without compromising the degree of optimality in terms of size of subset and corresponding dependency degree. Moreover, they improved algorithm which does not require any random parameter assumption and produce the same result every time. They also stressed the fact that two critical aspects of feature selection problem are the degree of optimality (in terms of subset size and corresponding dependency degree) and time required to achieve this optimality. Existing methods achieved success in either of these aspects such as Quick Reduct and EBR methods finds reduct in less time but not guaranteed to find a minimal subset [25], [30], [40].

In this part, we presented some of works done using FA in feature selections/reductions. FA proved to be suitable algorithms that have ability of dealing with multimodality. In other words, FA automatically subdivides populations into subgroups, and each group can swarm around each mode or local optimum. This automatic subdivision ability makes it particularly suitable for highly nonlinear, multimodal optimisation problems. In addition, the parameters in FA can be tuned to control the randomness as iterations proceed, so that convergence can also be sped up by tuning these parameters. This successfully led FA flexible to deal with continuous problems, clustering and classifications, and combinatorial optimisation as well. Nevertheless, FA has some advantages commonly highlighted in the research works. One of drawback of FA is hardly to fine balance between the right amount of exploration and the right degree of exploitation. This is because too much exploration increases the probability of finding the global optimality, while strong exploitation tends to make the algorithm being trapped in a local optimum. Since FA is a meta-heuristic, performance of FA highly dependent on fine-tuning the right amount of randomness and balancing local search and global search. Based on advantage and disadvantage presented, FA still need more research to be done especially on how to balance exploitation and exploration with fine tuning the randomness method in order to produce a good result and performance of the algorithms.

3.2.7 Summary

In this section, we presented six (6) popular natural inspired algorithms and their application in feature selections/reductions. The most common used algorithm in in feature selections/reductions were PSO and ACO because of its popularity and one of the earliest swarm algorithms been introduced. However, more algorithm has been developed and inspired from nature such as AFSA, ABC, FA and the latest which is BA. All this algorithm has demonstrated their capability and suitability in feature selection. Moreover, there is an effort has been done to hybrid the algorithm with other methods such as rough set and genetic algorithm to solve very specific problem and the result it quite promising. However, it is very important to mention that there is no perfect algorithms for all problem since each algorithms has pros and cons to solve very specific optimization problems. Some algorithms still suffered to balance the amount of exploration and exploitation in order to produce an optimum result. This kind of problem can be found in FA. Also in some algorithms while producing very promising result in term of reduction still lacking in time processing speed. All this crucial factors and mechanism of trade-off between factors need to be considered in the feature selections/reductions process.

4 CONCLUSION

Nature-inspired meta-heuristic algorithms have gained popularity because of their ability of dealing with nonlinear global optimisation problems. We have briefly reviewed the six (6) popular natural inspired algorithms and their application in feature selections/reductions. There is no doubt that all algorithms reviewed will be applied in solving more challenging problems in the near future, and its literature will continue to expand.

On the other hand, we have also highlighted the advantages and disadvantages of all algorithms specifically in feature selections/reductions. It is important to point out that all algorithms attempt to produce a good optimality feature without compromising accuracy in the data classification. Finally to conclude that there are two types of optimality which are:-

i. Optimality that concerns that for a given algorithm what best types of problems it can solve which is easy to be done because algorithm can be tested by a wide range of problems and then select the best type of the problems the algorithm of interest can solve.
ii. Optimality that concerns a given problem and try to find the best algorithm for efficient solutions. It can be done by comparing a set of algorithms to solve the same optimisation problem and hope to find the best algorithm(s). However in real world, there may be no such algorithm at all or all test algorithms may not produce a sufficient/good result.

The theoretical understanding of meta-heuristics is still lacking behind. In fact, there is more studies are highly needed in the area of feature selections/reductions using meta-heuristic algorithms in the future.

ACKNOWLEDGMENT

The authors wish to thank Universiti Utara Malaysia (UUM). This work was supported in part by a grant from UUM.

REFERENCES

[1] M. Dorigo and T. Stützle, "Ant Colony Optimization," *IEEE Comput. Intell. Mag.*, vol. 1, pp. 28–39, 2004.

[2] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proc. ICNN'95 - Int. Conf. Neural Networks*, vol. 4, pp. 1942–1948, 1995.

[3] Y. Cai, "Artificial Fish School Algorithm Applied in a Combinatorial Optimization Problem," no. November, pp. 37–43, 2010.

[4] H. A. Abbass, "MBO: marriage in honey bees optimization-a Haplometrosis polygynous swarming approach," *Proc. 2001 Congr. Evol. Comput. (IEEE Cat. No.01TH8546)*, vol. 1, 2001.

[5] D. Karaboga, "AN IDEA BASED ON HONEY BEE," 2005.

[6] P. Lucic and D. Teodorovic, "Transportation modeling: an artificial life approach," in *14th IEEE International Conference on Tools with Artificial Intelligence, 2002. (ICTAI 2002). Proceedings.*, 2002, pp. 216–223.

[7] D. T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, and M. Zaidi, "The Bees Algorithm – A Novel Tool for Complex Optimisation Problems," in *Intelligent Production Machines and Systems - 2nd I*PROMS Virtual International Conference*, 2006, pp. 454–459.

[8] K. M. Passino, "Biomimicry of Bacterial Foraging for Distributed Optimization and Control," *Control Systems IEEE*, vol. 22, pp. 52–67, 2002.

[9] K. N. Krishnanand and D. Ghose, "Detection of multiple source locations using a glowworm metaphor with applications to collective robotics," in *Proceedings - 2005 IEEE Swarm Intelligence Symposium, SIS 2005*, 2005, vol. 2005, pp. 87–94.

[10] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5792 LNCS, pp. 169–178.

[11] T. C. Havens, C. J. Spain, N. G. Salmon, and J. M. Keller, "Roach infestation optimization," in *2008 IEEE Swarm Intelligence Symposium, SIS 2008*, 2008.

[12] X.-S. Yang and A. H. Gandomi, "Bat algorithm: a novel approach for global engineering optimization," *Eng. Comput.*, vol. 29, no. 5, pp. 464–483, Jul. 2012.

[13] X. S. Yang and S. Deb, "Cuckoo search via Levy flights," in *2009 World Congress on Nature and Biologically Inspired Computing, NABIC 2009 - Proceedings*, 2009, pp. 210–214.

[14] C. Selvaraj, S. K. R, and M. Karnan, "A Survey on Application of Bio-Inspired Algorithms," vol. 5, no. 1, pp. 366–370, 2014.

[15] K. E. Parsopoulos and M. N. Vrahatis, "Recent approaches to global optimization problems through Particle Swarm Optimization," *Artif. Intell.*, vol. 1, pp. 235–306, 2002.

[16] A. Abraham, H. Guo, and H. Liu, "Swarm intelligence: Foundations, perspectives and applications," *Stud. Comput. Intell.*, vol. 26, pp. 3–25, 2006.

[17] M. Dorigo and G. Di Caro, "Ant colony optimization: a new meta-heuristic," *Proc. 1999 Congr. Evol. Comput. (Cat. No. 99TH8406)*, vol. 2, 1999.

[18] X. Li, "A New Intelligent Optimization-artificial fish swarm algorithm," Zhejiang University, 2003.

[19] A. Singh, "An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem," *Applied Soft Computing*, vol. 9. pp. 625–631, 2009.

[20] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, pp. 108–132, 2009.

[21] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Glob. Optim.*, vol. 39, pp. 459–471, 2007.

[22] X. S. Yang, *Nature-Inspired Metaheuristic Algorithms Second Edition*. 2008.

[23] S. K. Pal, C. . Rai, and A. P. Singh, "Comparative Study of Firefly Algorithm and Particle Swarm Optimization for Noisy Non-Linear Optimization Problems," *Int. J. Intell. Syst. Appl.*, vol. 4, no. 10, pp. 50–57, Sep. 2012.

[24] X. S. Yang, "A new metaheuristic Bat-inspired Algorithm," in *Studies in Computational Intelligence*, 2010, vol. 284, pp. 65–74.

[25] B. Yue, W. Yao, A. Abraham, and H. Liu, "A New Rough Set Reduct Algorithm Based on Particle Swarm Optimization," pp. 397–406, 2007.

[26] D. Ye, Z. Chen, and J. Liao, "A New Algorithm for Minimum Attribute Reduction Based on Binary Particle Swarm Optimization with Vaccination," *PAKDD*, pp. 1029–1036, 2007.

[27] G. Selvaraj and S. Janakiraman, "Improved Feature Selection Based on Particle Swarm Optimization for Liver Disease Diagnosis," *SEMCCO*, pp. 214–225, 2013.

[28] D. P. Rini, S. M. Shamsuddin, and S. S. Yuhaniz, "Particle Swarm Optimization : Technique , System and Challenges," *Int. J. Comput. Appl.*, vol. 14, no. 1, pp. 19–27, 2011.

[29] H. Yu, G. Wang, and F. Lan, *Rough Sets and Current Trends in Computing*, vol. 5306. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 242–251.

[30] R. Jensen and Q. Shen, "Finding rough set reducts with ant colony optimization," *Proc. 2003 UK Work.*, vol. 1, no. 2, pp. 15–22, 2003.

[31] Y. Jiang and Y. Liu, “An Attribute Reduction Method Based on Ant Colony Optimization,” in *2006 6th World Congress on Intelligent Control and Automation*, 2006, vol. 1, pp. 3542–3546.

[32] M. Mafarja and D. Eleyan, “Ant Colony Optimization based Feature Selection in Rough Set Theory,” vol. 1, no. 2, pp. 244–247, 2013.

[33] C. Bae, W. C. Yeh, Y. Y. Chung, and S. L. Liu, “Feature selection with Intelligent Dynamic Swarm and rough set,” *Expert Syst. Appl.*, vol. 37, pp. 7026–7032, 2010.

[34] Y. Chen, D. Miao, and R. Wang, “A rough set approach to feature selection based on ant colony optimization,” *Pattern Recognit. Lett.*, vol. 31, pp. 226–233, 2010.

[35] H. Fan and Y. Zhong, “A Rough Set Approach to Feature Selection Based on Wasp Swarm Optimization ⋆ Rough set theory,” vol. 3, no. February, pp. 1037–1045, 2012.

[36] K. R. Robbins, W. Zhang, J. K. Bertrand, and R. Rekaya, “The ant colony algorithm for feature selection in high-dimension gene expression data for disease classification,” *Math. Med. Biol.*, vol. 24, pp. 413–426, 2007.

[37] L. Ke, Z. Feng, and Z. Ren, “An efficient ant colony optimization approach to attribute reduction in rough set theory,” *Pattern Recognit. Lett.*, vol. 29, pp. 1351–1357, 2008.

[38] R. K. Sivagaminathan and S. Ramakrishnan, “A hybrid approach for feature subset selection using neural networks and ant colony optimization,” *Expert Syst. Appl.*, vol. 33, pp. 49–60, 2007.

[39] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, “Text feature selection using ant colony optimization,” *Expert Syst. Appl.*, vol. 36, pp. 6843–6853, 2009.

[40] N. Suguna and K. Thanushkodi, “A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony,” vol. 2, no. 6, pp. 49–54, 2010.

[41] N. Suguna, K. G. Thanushkodi, and T. Nadu, “An Independent Rough Set Approach Hybrid with Artificial Bee Colony Algorithm for Dimensionality Reduction,” vol. 8, no. 3, pp. 261–266, 2011.

[42] A. Chouchoulas and Q. Shen, “Rough set-aided keyword reduction for text categorization,” *Applied Artificial Intelligence*, vol. 15. pp. 843–873, 2001.

[43] P. Shunmugapriya and S. Kanmani, “Artificial Bee Colony Approach for Optimizing Feature Selection,” vol. 9, no. 3, pp. 432–438, 2012.

[44] P. Shunmugapriya, S. Kanmani, S. Devipriya, J. Archana, and J. Pushpa, “Investigation on the Effects of ACO Parameters for Feature Selection and Classification,” vol. 108, 2012.

[45] M. Schiezaro and H. Pedrini, “Data feature selection based on Artificial Bee Colony algorithm,” *EURASIP J. Image Video Process.*, vol. 2013, p. 47, 2013.

[46] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, “A comprehensive survey: artificial bee colony (ABC) algorithm and applications,” *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 21–57, Mar. 2012.

[47] M. Zhang, C. Shao, F. Li, Y. Gan, and J. Sun, “Evolving Neural Network Classifiers and Feature Subset Using Artificial Fish Swarm,” in *2006 IEEE International Conference on Mechatronics and Automation, ICMA 2006*, 2006, vol. 2006, pp. 1598–1602.

[48] G. Wang and D. Dai, “Network Intrusion Detection Based on the Improved Artificial Fish Swarm Algorithm,” *J. Comput.*, vol. 8, no. 11, pp. 2990–2997, Nov. 2013.

[49] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, “Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications,” *Artif. Intell. Rev.*, May 2012.

[50] H. Chen, S. Wang, J. Li, and Y. Li, “A Hybrid of Artificial Fish Swarm Algorithm and Particle Swarm Optimization for Feedforward Neural Network Training,” *Proc. Intell. Syst. Knowl. Eng.*, pp. 2–5, 2007.

[51] X. Zhang, F. Hu, J. Tang, C. Zou, and L. Zhao, “A kind of Composite Shuffled Frog Leaping Algorithm,” in *Proceedings - 2010 6th International Conference on Natural Computation, ICNC 2010*, 2010, vol. 5, pp. 2232–2235.

[52] D. Yazdani, S. Golyari, and M. R. Meybodi, “A new hybrid algorithm for optimization based on artificial fish swarm algorithm and cellular learning automata,” *2010 5th Int. Symp. Telecommun.*, pp. 932–937, 2010.

[53] R. Y. M. Nakamura, L. a. M. Pereira, K. a. Costa, D. Rodrigues, J. P. Papa, and X.-S. Yang, “BBA: A Binary Bat Algorithm for Feature Selection,” *2012 25th SIBGRAPI Conf. Graph. Patterns Images*, pp. 291–297, Aug. 2012.

[54] P. W. Tsai, J. S. Pan, B. Y. Liao, M. J. Tsai, and V. Istanda, “Bat Algorithm Inspired Algorithm for Solving Numerical Optimization Problems,” in *Applied Mechanics and Materials*, 2012, vol. 148–149, pp. 134–137.

[55] T. C. Bora, L. dos S. Coelho, and L. Lebensztajn, “Bat-Inspired Optimization Approach for the Brushless DC Wheel Motor Problem,” *IEEE Trans. Magn.*, vol. 48, no. 2, pp. 947–950, Feb. 2012.

[56] H. Banati and M. Bajaj, “Fire Fly Based Feature Selection Approach,” vol. 8, no. 4, pp. 473–480, 2011.

[57] H. Liu, A. Abraham, and Y. Li, “Nature Inspired Population-Based Heuristics for Rough Set Reduction,” in *Rough Set Theory: A True Landmark in Data Analysis*, vol. 174, A. Abraham, R. Falcón, and R. Bello, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 261–278.