International Journal of Future Generation Communication and Networking
Vol. 13, No. 2, 2020 pp.854-865

A Comparative Study on Bio Inspired Algorithms for Feature Selection In Machine Learning

J. A. Esther Rani[1], Dr. E. Kirubakaran[2]& Dr. B. Smitha Evelin Zoraida[3]

1. Research Scholar, Department of Computer Science,
Bharathidasan University,
Tiruchirappalli-620023
Tamilnadu,
South India

2. Professor, Department of Computer Science & Engineering
Karunya Institute of Technology and Sciences,
Coimbatore -641114
Tamilnadu,
South India

3. Assistant Professor, School of Computer Science, Engineering and Applications
Bharathidasan University,
Tiruchirappalli-620023
Tamilnadu,
South India

Abstract

Mother Nature is a great inspirer of this universe. Nature always finds a way to exist and motivates us to solve tough and multifarious problems in the realm of Information Technology and Computer Science. Nature solves its problems on its own and this is the main key for exploring the nature inspired algorithms. Here the optimal solution is derived by maintaining a balance between all the components. These algorithms are imitated by the way the nature solves its own problems and renders a feasible solution. The nature inspired algorithms need to be hybridized not only to get the best possible solution but also to solve the same by maintaining the perfect sense of balance among its components. This motivation has led to pursue research work on nature inspired computing. This work presents a comparison of various nature inspired algorithms which can be used for feature selection. It can be extended to all the domains including medical, web, user profiling and so on.

Keywords*: Nature Inspired Algorithm, Optimization algorithms, Feature selection, Hybridization*

Introduction

Optimization is the one which finds the best feasible solution which commonly exist in scores of disciplines. In nature a deterministic approach or stochastic approach is present and the same can be used in Optimization algorithms. Deterministic approach requires huge computational labour, which fails as the magnitude of the problem increases. This leads to the insight for using nature inspired stochastic approach for optimization algorithms. Meta heuristics provides adequately a good solution for optimization problem and it uses techniques involving randomization and various local search algorithms to find a solution for a given problem as it performs in Evolutionary algorithms, Swarm based algorithms and Simulated annealing search.

Nature as a Problem Solver

The nature inspired algorithms as its name suggests gets its inspiration from nature. They apply simple conditions and rules to explain and solve complex relationships without even having detailed facts of the search space. The best optimal solution provider is nature. Physical phenomena like formation of river, destruction through forest fire, formation of clouds, rain, thunder etc. are examples of nature giving optimal solutions for problems that exist in nature. A straight forward

ISSN: 2233-7857 IJFGCN

association can be made possible between nature and technology. Nature inspired computing provides optimal solution in computing which includes areas like network security, robotics, bioinformatics, Big data etc. This paper compares various Nature Inspired optimization Algorithms for feature selection and discusses their respective application areas.

About Evolutionary Algorithms:
Evolutionary computation in short known as EC [1] is an idea used in the knowledge engineering area which targets to take advantage from combined prodigy in a customizable large group of problem solvers who employ the repetitive progress which includes different phases like growth, development, reproduction, selection, and survival that are included within a given population. Prominent successful algorithms are algorithm based on Genetics. The condition they use is best to survive and belong to population based stochastic search algorithms.

Intelligent Water Drops Algorithm (IWD)
A novel population based method proposed by Hamed Shah-hosseini in the year 2007 [12] is Intelligent water drops algorithm. It is inspired with the processes involved in the formation of river systems constituting the steps and the subsequent reactions that happen between the water drops in the waterway and the changes that occur in the surroundings as the river is formed. After observing the behaviour of the drops, an artificial water drop algorithm is formed that inherits some of the significant properties of the natural water drop. The Intelligent Water Drop algorithm poses two significant features: 1.Soil (IWD) (i.e.) the quantity of the soil it carries currently. 2. Velocity (IWD) (i.e.) the velocity with which it moves currently. The problem under consideration will be the location and the surroundings in which the water flows. The algorithm advances in distinct countable-length steps. From its present position to its subsequent position, the IWD velocity increases. The total velocity increased is calculated as discretely relative to the inverse of the soil that lies in-between the present location and the subsequent locations. Some amount of the soil that lies between the paths is removed when both the locations are connected. The removed soil is included to the IWD. It is indirectly relative to the time required for the IWD to move from its existing position to the subsequent location. The total time for this process is estimated using the velocity of the IWD and the distance between the existing and the subsequent locations. IWD takes the tracks with light soil on its beds to the paths with heavy soils on its beds. The chance of choosing the subsequent track is always indirectly proportional to the soils available in the track.
Disadvantage of IWD:
Occasionally IWD will not lead to the feasible solution. Random probability is the reason behind this [24].

Swarm intelligence
Next we will look into Swarm Intelligence (Kennedy and Eberhart, 2001 [7]) algorithm that is very widely used by researchers in analysing and implementing adaptive systems. The dissimilarity between the Evolutionary algorithms and the Swarm Intelligence is that the former takes on genetic change of organisms for their survival and latter takes on combined social behaviour of the organisms. The tool used in finding the optimal solutions is the behaviour pattern of the Swarm. Here the word "swarm" refers to the uneven activities (i.e.) the locomotion of the particles in the problem domain. An SI algorithm takes motivation from the group behaviour of animals' specifically self organized behavioural patterns deployed by them in the search of food.

The principles involved in the SI are basically categorised into five basic principles.

The Principles behind Swarm Intelligence: SI can be explained with the following five basic principles.

No.1: Awareness
Every member in the swarm should be aware of its environment and abilities. It must be able to perform time and space calculation.

ISSN: 2233-7857 IJFGCN

No.2: Autonomy
Every member in the swarm must function as a self-directed master. This is necessary to self-regulate the allocation of work.

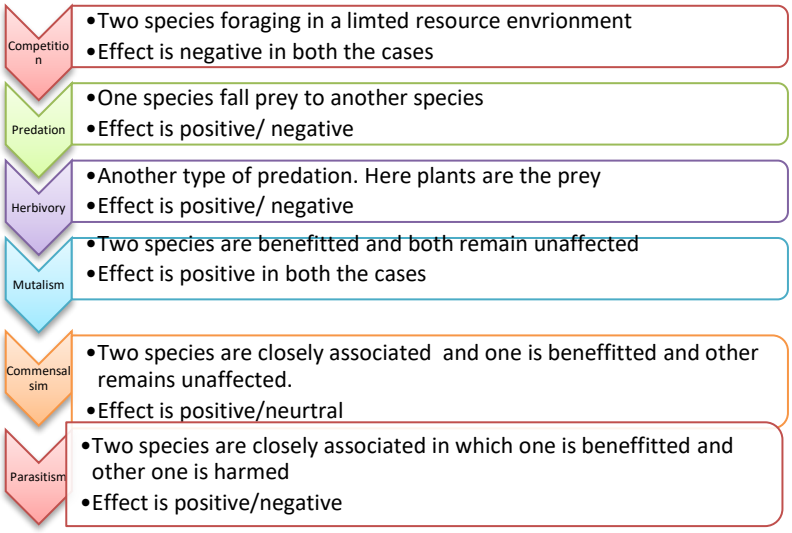
No.3: Solidarity
Every member in the swarm must help in solidarity. When task in hand is completed every member must search for a new task.

No.4: Expandability
The entire group (swarm) must adapt expansion where members are dynamically grouped.

No.5: Resiliency
The system must be adaptable (i.e.) self-healing when any member in the system is removed; the remaining members should carry out the unfinished tasks.

Environmental Science:

The environmental science that exist in nature reveals the techniques for addressing complex problems in engineering and Technology. The ecological system includes all organisms along with the abiotic factors such as non living chemical and physical parts such as air, soil, water etc. The species present in ecological system interact with each other. The interaction can exist in two modes, one is cooperative which represents sociality and division of labour and the other one will be competitive. Also there are two types of interaction that exist in the environmental system. They are the interspecies interaction and intra species interaction. The interaction between inter species can be categorized into three main types with respect to the outcome of interaction. If the interaction is positive it is termed as mutualism, if it is negative it is parasitism and it is commensalism when the interaction is neutral. The following chart reveals the various types of interspecies interaction in the environmental system.



This reveals the different ways and means in which social evolution and the social foraging happens. Biogeography studies the distribution of species over a particular time and space.

Human Innate Immune System:
The human body without our knowledge is constantly at war. We are surrounded by harmful bacteria, viruses, parasites and toxins. It affects the health of the human beings. To overcome this, the body uses the natural immune system to fight against anything that may cause a threat to the human body. The natural immune systems are the parts in the immune system that works automatically, no matter what and how the damage is caused. They all aim at protecting the body without the requiring a lot

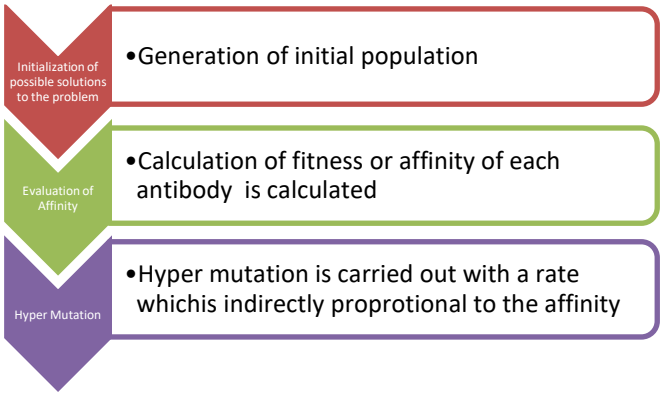
ISSN: 2233-7857 IJFGCN

of preparation. They never cease to work and they start invading the offender even when they have not seen the offending invader earlier.

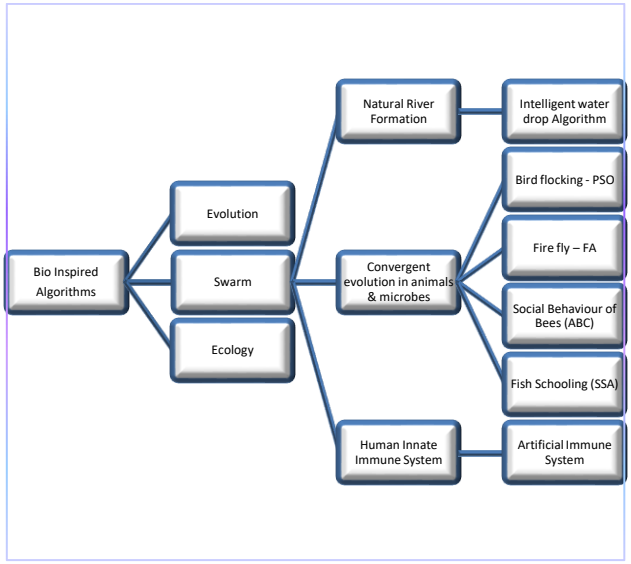
Artificial Immune System (AIS) Algorithm

This algorithm AIS was developed by Dasgupta, in 1999 [14]. AIS algorithm takes inspiration on how the human immune system functions. It has the features of self healing such as self-organization, learning from the current situation, adaptation, memory, robustness and scalability [26]. AIS emulates the human immune system

The following are the process involved in AIS are given below:



A Comparison of Nature inspired Optimization algorithms for Feature Selection



Particle Swarm Optimization

Particle swarm optimization (PSO) was developed by Kennedy and Eberhart in 1995[8]. It models the social behaviour of animals like an army of ants, flight of birds, school of fishes and pack of animals. The technique this algorithm uses is a population based stochastic optimization technique. PSO has several unique features such as simplicity, specific searching techniques, computational intelligence and easy implementation. It is widely used to optimize problems involved in engineering and technology.

In PSO, the letter 'P' represents hundreds or thousands of particles which are of very less mass and volume. By communicating with one another they search and find the optimum value. Each particle 'p'is defined by two state vectors:

(i) By position x_i^k and (ii) By velocity v_i^k. These state vectors are updated as given below [27]:

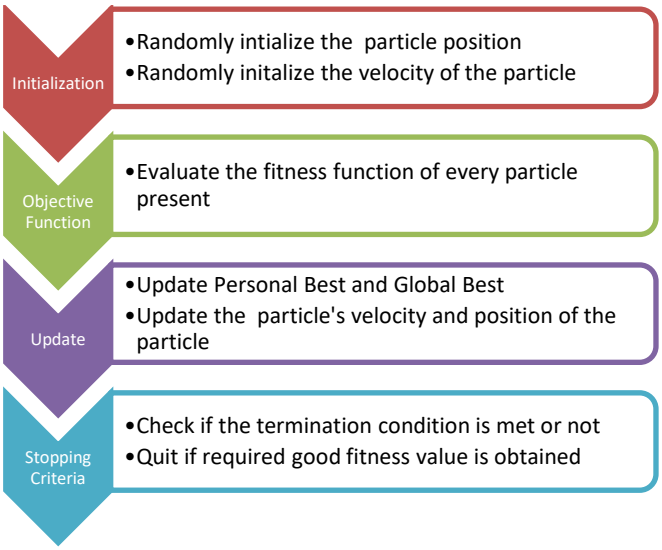
$$v_i^k = wv_i^k + c_1 r_1 (p_i^k - x_i^k) + c_2 r_2 (p_g^k - x_i^k) \quad (1)$$

ISSN: 2233-7857 IJFGCN

$x_i^k = x_i^k + v_i^k$ (2)

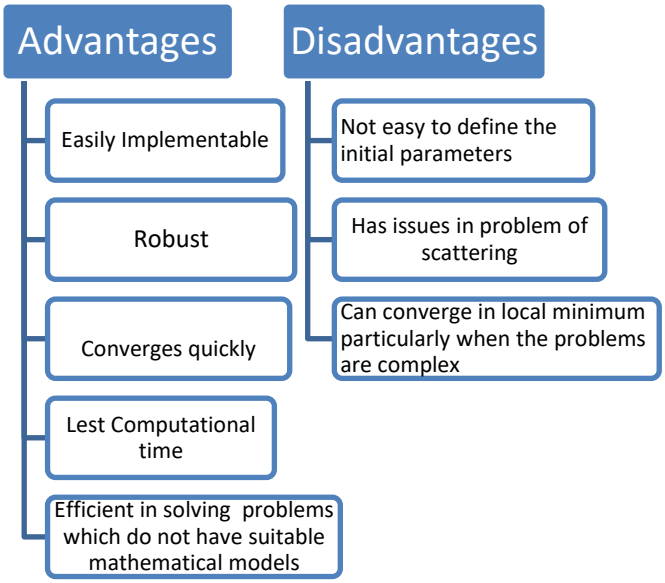
Where, x_i^k refers Particle position
v_i^k refers Particle velocity
p_i^k refers best "remembered" position
c_1, refers cognitive parameter
c_2 refers social parameters
r1, r2 refers random values ranges from 0 to 1

Process in PSO algorithm:



Advantages and Disadvantages of PSO:

The following table evaluates on the advantages and disadvantages of Particle Swarm Optimization algorithm:



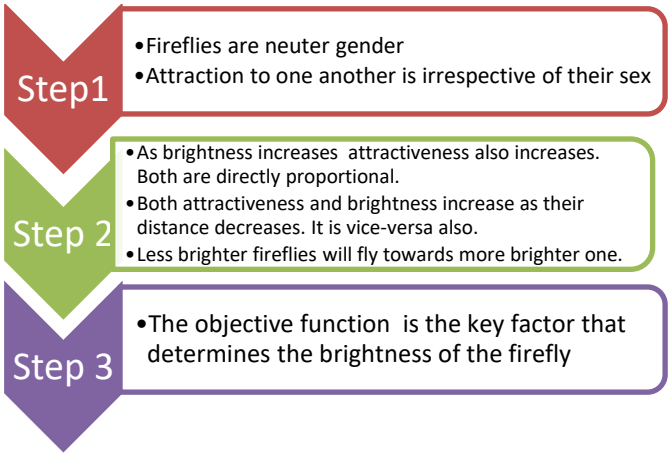
Firefly algorithm

Firefly algorithm is another algorithm which belongs to the nature inspired algorithm proposed by Yang [15]. It is classified as meta heuristic and comes from the basis of swarm intelligence. The author drew the inspiration from the twinkling behaviour of fireflies which formed the key to solve

ISSN: 2233-7857 IJFGCN

constrained optimization problems. The group of fireflies perform distinguishing luminary flashing actions such as to attract the partners, to establish communication, and to warn the group about the risk of predators. The Agents communicate with one another through bioluminescent radiant which guides everyone to explore more effectively and successfully than in randowm search with standard distribution.

Steps Involved in Firefly algorithm:



Advantages:

- Effectiveness of the algorithm is more in multi objective optimization.
- Fast in solving optimization and engineering problems

Disadvantages:

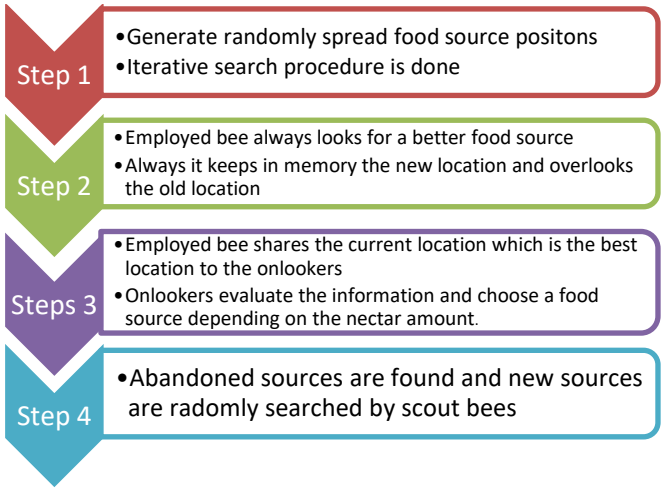
- Slow convergence speed
- More likely to be trapped in local optimum

Artificial Bee Colony

Artificial Bee Colony was developed by Karaboga and Basturk [10] modelling the nature inspired behaviour of honey bee swarm in foraging. In ABC algorithm, the colony of artificial bees is classified into three categories: one is employed bees, onlooker bees and scout bees. The first category is the employed bees which uses their memory to search around for the sources of food. It then shares the information about the sources of food to the onlooker bees. The second type of bee (i.e.) the onlooker bees choose the food source that is best from all the places found by the employed bees. Fitness which is higher will have greater chances to be chosen by the onlooker bees than the lower one.

The third kind of bee is translated from few employed bees to the scout bees which carry out new indiscriminate search for discovering new food sources. The location of the food source denotes an optimal solution to the optimization problem and fitness or the quality is measured with the nectar amount of the food source. The number of the employed bees in the colony is equal to the number of feasible solutions in the population.

ISSN: 2233-7857 IJFGCN



An employed bee alters the position (solution) in its memory according to the visual information on the neighbourhood and evaluates the nectar amount which is the fitness value of the current location (latest solution). If the fitness function is greater than the previous one than the new location will be memorized by the employee bee and the old location will be forgotten by it. When the entire searching process is complete the employed bees share their information (fitness value) and the position to the onlookers. In the reproduction phase the onlookers chooses the food source according to the probability value. The final phase is the selection phase which involves replacement of bees. In this phase it is decided whether the food source is abandoned and it is found that it cannot be continued through a finite number of cycles. The most important parameter is the predetermined number of cycles in the Artificial Bee Colony.

Advantages of Artificial Bee Colony:

- Simple and Robust
- Can find local solutions quickly
- Easy to implement

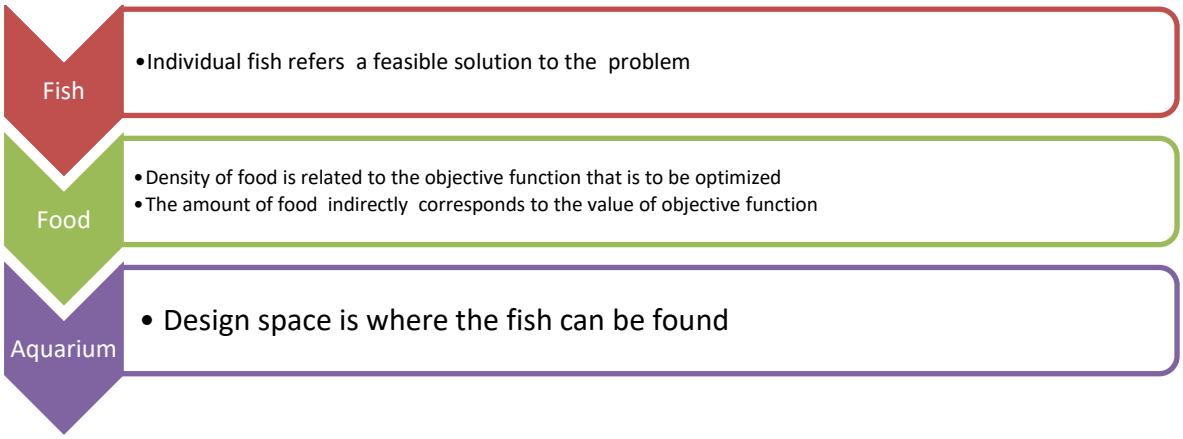
Disadvantages of Artificial Bee Colony:

- Very slow in solving problems involving sequential processing
- More number of objective function evaluation
- Secondary information is neglected

Fish Swarm Optimization Algorithm

The fish swarm optimization algorithm (FSOA) derives its inspiration from the shoal of fish that exist in ocean. The inspiration is drawn from the shoaling behaviour of fish. The technique used here is the evolutionary computation and it is a population based stochastic algorithm developed by Li et al. [11] in 2002.

The following features have to be considered:



ISSN: 2233-7857 IJFGCN

Fish swarm algorithms takes three different types of behaviours: [28]
(i) Searching for food by performing random search to reduce the objective function (i.e.) the food satisfaction
(ii) Swarming in response tries to satisfy the food intake needs and also aims to attract new swarm members.
(iii) Following helps the neighbouring individual to reach the food. The fishes will follow where there is greater amount of food thereby increasing the probability of achieving a successful result

Advantages:
- It has immense potential to prevent local minimums in a way to reach global optimization.
- High Accuracy is achieved
- Flexibility is higher

Disadvantages:
- Time consuming
- Advanced convergence

Salp Swarm Algorithm

Salp is an aquatic being which pertains to the lineage of Salpidae. It has an outline like a barrel which resembles to jellyfish. It has a cylinder shaped structure with openings at both the end. The openings help to propel water through their gelatinous bodies, to swim and for taking food [29]. The figure of a salp is shown below:

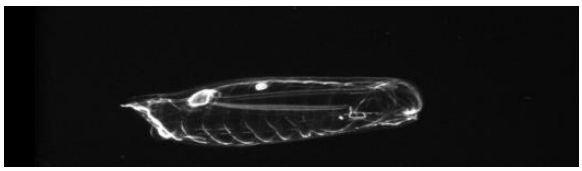


Fig a. A Salp

The schoaling behaviour in the deep oceans often forms a chain like structure which is known as salp chain. The rationale for this formation is not yet known but the scientist thinks it helps in better locomotion and for searching of food [1].

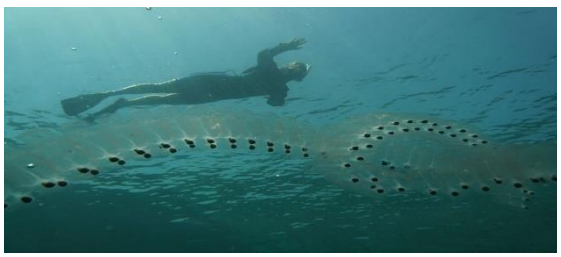
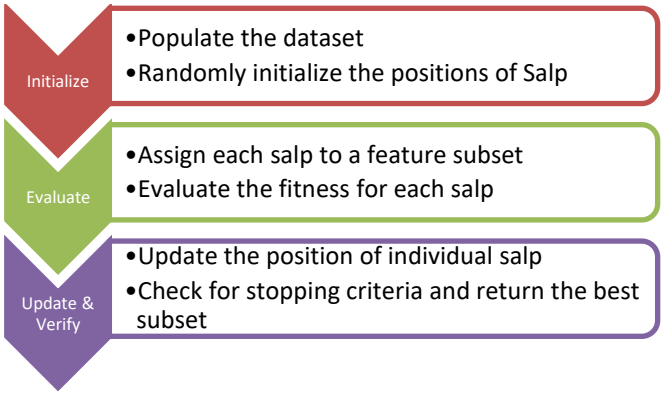


Fig b. Swarm of salps (salp chain)

The steps involved in Salp Swarm Algorithm

ISSN: 2233-7857 IJFGCN



To implement Salp Swarm algorithm the swarm is divided into two. One will be the head or the chief and the left behind salps in the chain are the followers. The leader directs the entire chain of salps and they will follow the leader either directly or indirectly. The representation of the position of salp will be represented in a two dimensional array. The target of the salp swarm chain is to identify the food source in the entire search space available. The following equation upgrades the leader position[30]:

$$x_j^1 = \begin{cases} F_j + c_1\left((ub_j - lb_j)c_2 + lb_j\right) & c_3 \geq 0 \\ F_j - c_1\left((ub_j - lb_j)c_2 + lb_j\right) & c_3 < 0 \end{cases}$$

where
x_j^1 refers the location of the head or leader
F_j ->location of the food source
ub_j->refers the upper bound
$lb_{j->}$ refers the lower bound
c_1, c_2, and c_3 are variables random values.
It clearly states that the location is updated only for the leader with reference to the position of the food source. The parameter c_1 which has the random value is the key coefficient. The main reason behind that is it provides equilibrium with exploration and exploitation. It is defined in the following equation [30]:

$$c_1 = 2e^{-\left(\frac{4l}{L}\right)^2}$$

where
l -> the present iteration
L ->denotes the upper limit of the iterations.
The parameters c_2 and c_3 are numbers with random values between 0 and 1. It is generated uniformly in the interval of [0, 1]. These parameters guides us to the next position which must be taken towards negative infinity or positive infinity. The location of the followers is upgraded by implementing Newton's law of motion:

$$x_j^i = \frac{1}{2}at^2 + v_0 t$$

where i≥2,
x_j^i->the location of i^{th} follower in the j^{th} dimension
t -> time taken
v_0 -> represents the initial speed
a is determined by the following equation:

ISSN: 2233-7857 IJFGCN

$$x_j^i = \frac{1}{2}\left(x_j^i + x_j^{i-1}\right)$$

where $i \geq 2$
x_j^i->the location of i^{th} follower in the j^{th} dimension

Thus the followers are updated and the slow and stepwise movement of them prevent the Salp Swarm algorithm to be stuck in local optima.
The key controlling parameter c1 is decreased accordingly over the cycle of iterations, which enables the salp swarm algorithm probe the search space first and then capitalize on it.

Advantages of Salp Swarm Algorithm:

- Hybridizing with similar algorithms will always lead to better results
- A good convergence takes place quickly
- Increased process in obtaining best solutions
- Very effective in finding solutions for different kinds of optimization problems
- Applicable for problems with wide search space
- Adaptability, robustness, and scalability
- Has high feasibility and efficiency in achieving global optimum.

Disadvantages of Salp Swarm Algorithm:

- Deteriorates with untimely convergence
- No hypothetical converging frame
- The distribution by probability alters by generations

VI. Conclusion

Bio inspired algorithms are creating a paradigm shift in the realm of Computer Science. These algorithms are inspired by nature and so their boundaries are boundless. So is the way of providing solutions to Computer Science problems. By hybridizing the algorithms we can proceed to the next generation modelling and computing. This work renders a summary of various algorithms to afford optimization and enhanced feature selection. The real time applications of nature inspired computing in past years are tremendous and they were the solution provider for various network related problem, data mining, web mining, biometrics, image processing, robotics, forecasting problems, etc.. There are arena where there are still notably demanding work for the researchers to provide solution by hybridizing the algorithms to offer perfect solution for new arising domains in engineering and technology. More specifically, there are innumerable ways and means to explore and hybridize a new approach or an algorithm. This paper proposes to synthesize two algorithms and hence we propose to hybrid Salp swarm algorithm with modified particle swarm optimization algorithm as a first step to provide a better extraction for feature selection in gene expression data.

REFERENCES

[1] Back, T. 1996: Evolutionary algorithms in theory and practice. Oxford University Press.

[2] J.H. Holland, Genetic algorithms and the optimal allocation of trials, SIAM J. Comput. 2 (2) (1973) 88–105

[3] Koza, John R. 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: The MIT Press.

[4] Beyer, H.G. and Schwefel, H.P. 2002: Evolution strategies. Natural Computing 1, 3–52.

ISSN: 2233-7857 IJFGCN

[5] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, Journal of Global Optimization 11 (1997) 341–359.

[6] UpekaPremaratne , JagathSamarabandu, and Tarlochan Sidhu, A New Biologically Inspired Optimization Algorithm‖, Fourth International Conference on Industrial and Information Systems, ICIIS 2009, 28-31 December 2009, Sri Lanka.

[7] Bonabeau, E., Dorigo, M. and Theraulaz, G.1999: Swarm intelligence. Oxford University Press

[8] Kennedy, J.; Eberhart, R. (1995). "Particle SwarmOptimization". Proceedings of IEEE International Conference on Neural Networks. IV. pp. 1942–1948.

[9] Dorigo, M., Maniezzo, V., &Colorni, A. (1996). Ant System: Optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics – Part B, 26, 29–41.

[10] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, Journal of Global Optimization 39 (2007) 459–471

[11] X. Li, Z. Shao, J. Qian, Anoptimizing method base on autonomous animates: fish- swarm algorithm, Systems Engineering Theory and Practice 22 (2002) 32–38.

[12] Shah_Hosseini, H. ShahidBeheshti Univ., Tehran Problem solving by intelligent water drops IEEE Congress on Evolutionary Computation, 2007. CEC2007.

[13] K. M. Passino, ―Biomimicry of bacterial foraging for distributed optimization and control,‖ IEEE Control Syst. Mag., vol. 22, no. 3, pp.52–67, Jun. 2002

[14] D. Dasgupta, Artificial Immune Systems and Their Applications, Springer, Berlin, 1999

[15] X.S Yang, ―Fire fly algorithm for multimodal optimization‖,in proceedings of the stochastic

[16] Algorithms. Foundations and Applications (SAGA 109) vol.5792of Lecture notes in Computer Sciences Springer, Oct.2009

[17] S.He,Q.H.Wu, Senior Member IEEE and J.R Saunders.A novel group search optimizer inspired by Animal Behavioral Ecology; 2006, IEEE Congress on Evolutionary Computation, 1272-1278

[18] Eusuff MM and K.E Lansey; Optimization of water distribution network design using SFLA(2003)

[19] Hanning Chen , Yunlong Zhu, Optimization based on symbiotic multi-species coevolution;journal on Applied Mathematics and Computation 205 (2008)

[20] A.R. Mehrabian, C. Lucas, A novel numerical optimization algorithm inspired from weed colonization, Ecological Informatics 1 (2006) 355–366.

[21] Simon, D., 2008. Biogeography-based optimization. IEEE Transactions on Evolutionary Computation. 12 (6), 702–713.

[22] P. A. Anderson and Q. Bone, "Communication between individuals in salp chains II. Physiology," Proceedings of the Royal Society of London B: Biological Sciences, vol. 210, pp. 559-574, 1980.

ISSN: 2233-7857 IJFGCN

[23] V. Andersen and P. Nival, "A model of the population dynamics of salps in coastal waters of the Ligurian Sea," Journal of plankton research, vol. 8, pp. 1091-1110,1986.

[24] N. Henschke, J. A. Smith, J. D. Everett, and I. M. Suthers, "Population drivers of a Thalia democratica swarm: insights from population modelling," Journal of Plankton Research, p. bv024, 2015.

[25] GunawanIndraMaryati,Gunawan" Comparison of Intelligent Water Drops and Backtrack Algorithm in Solving Travelling Salesman Problem" Information Systems International Conference (ISICO), 2 – 4 December 2013

[26] Mohamed A. Mohamed, Ahmed A. ZakiDiab, HegazyRezk. "Partial shading mitigation of PV systems via different meta-heuristic techniques", Renewable Energy, 2019

[27] J. Timmis, A. Hone, T. Stibor, and E. Clark, "Theoretical advances in artificial immune systems," Theoretical Computer Science, vol. 403, no. 1, pp. 11–32, 2008.View at: Publisher Site | GoogleScholar | ZentralblattMATH | MathSciNet

[28] A. Hidaka, and T. Kurita,"Non-Neighboring Rectangular Feature Selection using Particle Swarm Optimization",*19th International Conference on Pattern Recognition*, Tampa, FL, USA, pp. 1-4,2008.

[29] "A Survey of Bio inspired Optimization Algorithms" Binitha S, S Siva Sathya,International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-2, May 2012

[30] Salp swarm algorithm: a comprehensive survey by LaithAbualigah, Mohammad Shehab, Mohammad Alshinwan1 &HamzehAlaboolSpringer-Verlag London Ltd., part of Springer Nature 2019

[31] S. Mirjalili et al., Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems, Advances in Engineering Software (2017),http://dx.doi.org/10.1016/j.advengsoft.2017.07.002 S. Mirjalili et al. / Advances in Engineering Software 000 (2017) 1–29 5

ISSN: 2233-7857 IJFGCN