Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)
IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

Evolutionary Computation based Feature Selection: A Survey

Suresh Dara
Department of CSE
B V Raju Institute of Technology
(UGC Autonomous)
Narsapur, Telangana, India-502313.
darasuresh@live.in

Mamidi Jagadeeshwara Reddy
Department of CSE
B V Raju Institute of Technology
(UGC Autonomous)
Narsapur, Telangana, India-502313.
jagadeesh.reddy173@gmail.com

Nageswara Rao Eluri
Department of Computer Science
King Khalid University
Abha, Kingdom Of Saudi arabia
eluri76@gmail.com

***Abstract*—In previous years, different Lateral thinking optimization techniques have been developed based on evolutionary computation. Many of these methods are inspired by spill out behaviors in nature. In this Paper, a new optimization algorithm based on the law of gravity and mass interactions named as Gravitational Search Algorithm (GSA) is discussed for solving feature selection. In GSA, the searcher agents are a collection of masses which will interact with each other based on the law of motion and Newtonian gravity which gives the binary evolutionary optimized high performance. The detailed feature selection has been discussed in this paper and The GSA method has been compared with some well-known optimized search methods such as GA (Genetic Algorithm), PSO(Particle Swarm Optimization).**

***Index Terms*—Gravitational search algorithm, Heuristic search algorithms, Law of Gravity, Optimization**

I. Introduction

According to Instant based learning, to find out nearby instances we need distance function which will be computed in terms of features. If the number of features are larger in number we will have the feature reduction problem. Features contain information about the target. Classification function defines number of functions as we think that numbers of features are more which gives the better information or better discrimination power or better classification power but this theory not always hold.

In Machine learning and Statistics, feature selection is also known as variable selection, attribute selection, feature subset selection.

Fig. 1 define the theory of the performance with increase in the number of variables when the training examples are fixed.

The reason for the above degradation of the graph is: Irrelevant features and Redundant features.

When the search space larger then the time taken to compute will be more and it depends on the algorithm. The too many dimensions will called as the Curse Of Dimensionality. To Overcome the Curse Of Dimensionality we use the two concepts: They are: 1. Feature Selection, 2. Feature Extraction.

Feature selection is the combination of the algorithms for taking the set of features which are subsets.

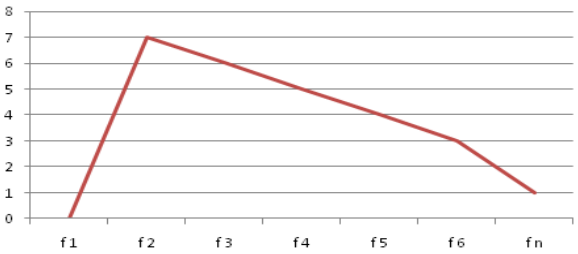

Fig. 1. Feature selection for 'N' Number of features

We initialize with 'N' Number of features

$$F = x_1, x_2, x_3, x_4, \ldots, x_n \tag{1}$$

Then we have to find the subsets of the features

$$F^1 \subset F = x_1^1, x_2^2, x_3^3, \ldots, x_m^1 \tag{2}$$

The number of possible subset for N Number of sets is 2^N, Which is not possible for n number of sets so we use algorithms for them.Feature selection(FS) is active in different fields such as data processing, data mining, machine learning, classification problems. Until now, large number of methods for FS have been introduced and reported in literatures. Based on the algorithm selection and the model building the methods are categorized into three groups: I. Filter method, II. Wrapper method, III. Embedded method or hybrid method [1].

The filter based method look at the input only which work using the dataset and use the statistical info of the data to select features which is unsupervised technique. For the given data, it selects the subset using search strategy which has the most information in it. Each generated subset is evaluated individually or isolatedly. The subset which with the best evaluation will be the best subset. The search will continue till some of the pre defined stopping situations are met. In the Filter method the feature doest depend on the classifiers. Better is used in the large scale datasets [2], [3]. Wrapper method evaluates using the learning algorithm which depends on classifiers to enlarge

the accuracy of the classification using the supervised technique. These methods assume the transition using the selected subset and learning algorithm model and preferred in the problems of the classification. Wrapper algorithms are same as the above unsupervised technique algorithm but they use the predefined mining algorithm in spite of an individual measure for the subset selection. Hybrid or Embedded approaches which takes the advantage of both the supervised and unsupervised technique which are filter and wrapper based techniques. These methods use the individual and independent measure to select the subsets for given the number of elements in a set or other grouping, as a property of that grouping and use the learning algorithm to select the best subset which is the final among the best subsets across given the number of elements in a set or other grouping, as a property of that grouping.There are studies that has given the introduction of the methods in the areas of the feature selection [1], [4], [5].

The determination of the non-consequent features is a complex issue. The problems of FS increases with increase in the features. Thus the search over the solution space is not practical. In certain categories, many methods of the feature selection are based on the search algorithm where innovative search algorithms are utilized. Sequential Forward Selection and Sequential Backward Elimination and many techniques are utilized for many number of years. Many recommended feature selection methods which are meta-heuristic search algorithm for the subset selection process. The essential optimization algorithms which have an efficient global search technique [2]. The advantages of EC over the methods of traditional,it is used in Feature selection for various domains areas including image analysis [6], face recognition [7], software fault [8]. It has been taken for datasets of biomedical [9] and shown reduction in features. To analyze the domains, researches have used methods like particle swarm optimization (PSO) [10], genetic algorithm(GA) [11], among others.

Optimization improves or maintains the classification accuracy and simplifies the classification complexity. There are approaches in the feature selection: 1. Forward selection, 2. Backward Elimination.

The Forward selection starts with no variables and adding them one by one, at every step adding which will not increase the error the most, till any addition which will not decreases the error.

The Backward Elimination which starts with the full model and removes one by one, at each step removes the predictors which will not necessary or does not increase the error till further removal significantly. Which is based on the validation set which is different from the training set.

II. Background

In this section, we discuss some of optimization methods like GA, PSO, and GSA.

A. GA

The Genetic Algorithm is search based optimization Technique [12], based on the principle of Natural Genetics and Natural Selection which finds the optimal solution to the problem which is based on the Survival of the fittest solution. Genetic Algorithm is not the traditional algorithm which is used to solve the clustering, classification and estimation instead Genetic Algorithm is technique or approach to find the final optimal solution where the Natural Genetics is used. Natural Genetics solves the crossover and mutation and produces the new children. Crossover may vary significantly during the beginning when the population is randomly initialized so that it has important effect moving along the chromosomes large area in space of problem. At the last run, not most chromosome have smaller structure it has minute effect, lesser movement is possible starting with the large probability and ending with the smaller one [13]. Mutation is going through the Genetic Algorithm different to the crossover which has small effect at the beginning and more effect at the end because initially the randomized population is given,changing here and there does not effect it significantly but after some iterations having smaller values it may have large impact on the solution set. The Process is repeated or iterated until the optimal values taken. Considering the parents to the childrens produces the same chromosomes which is crossover assuming it produces the better optimal solution, The mutation is the process where the children chromosome have additional feature which doest there in the parent chromosome. Such that two crossover and mutation are performed to generalize the new solution.

The new solution is the one which most fitted according to the objective function taken as compared to the parent solution. The best fits only are considered and taken to the next level Each individual are assigned a fitness value based on the Objective function value and fitter individual are given a highest to mate and yield more fitter individual. Here, fitness function takes the condition solution to the problem as input and produces the output as the fitness solution in response to problem in deliberation which uses the stochastic operations. At last it uses the selection to select the best fitted optimal value. Genetic Algorithm is theoretically non-zero probability any state can be occupies the chromosome [14]. But in Practical it is not possible because number of steps are required. Selection uses the Tournament selection, Any of the selection can be used but it should contain mainly Elitism strategy.

B. PSO

In the field of Computers research, was originally designed and developed by Eberhart and Kennedy [15].A PSO is a procedure which is raising from external creative impulse by Natural selection belongs to the more number of development algorithms. Which represent the behavior of flocking of birds. Initiating from random particle set

generally named as Solutions. The Particle Swarm Algorithm improves using the fitness function to the set. Here, the development happens when moving the set of particles around the N Dimensional area(space) using set of expression which are simple which gives some inter particle communications. The mathematical expressions are in simplest form, the movement of particle towards own best position suggesting the best position of the swarm till then, Different rules used by the variants so far. It is generally utilized as a technique for optimization, but it has deep knowledge in animation field where developed a system of the particle as a individual working to form the fuzzy object like Cloud. The thought was to start the set of points for giving the velocity to them. Using the given vectors, the position of each particle changes repeatedly while random numbers will adjust the vectors. Reynolds [16] summed up the notion of the communication between the objects to Reeves [17] system of particle to give the algorithm for flocking where the particle should follow rules of flocking such as matching the velocities of each other.

That system helps for modeling difficult vast behaviors group in simplest way. In the process of improvement of their paradigm, Millonas [18] they discussed five principles of swarm intelligence. The proximity principle is the first one where the population should able to do the small area and time calculations. The Quality principle is the second one where the population should be able to react quickly to quality particles in the system. The diverse response is the third principle where the particle should not do its work along small channels. The stability principle is the fourth one where the changes in the environment should not affect the particle behavior. The adaptability principle is the fifth one when the calculation price worth appears particle should be able to change the mode of the behavior. They compromisingly call less mass and less volume particle members such that the concepts such as acceleration and velocity more benefit. Thus, the particle swarm optimization term was given [19].

PSO is an easily understood algorithm which seems to be effective for optimizing a broad area of functions. We will not view this as a high-level, We view this as a mid-level or derived from the biological frame algorithm, Which take up the area in nature of the evolutionary search, which takes long period of time, takeup the order of the milliseconds. The optimization of social happens in the ordinary experience frame of time.

Conceptually, the PSO lies between the GA and EP and it is mainly occur on debatable thing, like EP (Evolutionary Algorithm and programming). The modification towards the xbest and the ybest by the PSO is approaching same to the crossover operation used by GA(Genetic Algorithms). Which use the process of fitness as all the Evolutionary algorithms [15].

Many feature selection algorithms have been implemented in past years by using meta-heuristics. PSO based feature selection algorithms proposed with different variations like using Binary PSO [20], an Elitist BPSO [21], Hamming Distance based BPSO [22] [23], Multi Objective BPSO [24] and Rough based hybrid BPSO [25].

C. GSA

In the Gravitational law of Newton, each and every particle attracts every another particle with gravitational force. The GF between two particles is directly proportional to the multiples of the masses and inversely proportional to the square of the distance between the particles.

In this section we will have a small explanation of GSA. In GSA the objects will be agents and the performance will be measured by their mass. All the objects attracts every other object by GF, and Gravitational force which action the movement of the objects globally to-wards other objects with heaviest mass. The heavier masses which gives best result of the problem.

In the Gravitational search algorithm, every agent has four Attributes: position of the agent, its inertial mass, active gravitational mass and passive gravitational mass. Position of the mass corresponds to the problem solution, its inertial and gravitational masses are determined by using the fitness function.

The Gravitational search algorithm be considered as an incommunicado system of masses. It is artificial world of masses which agreeing the Newtonian Law of motion and gravitational. In Other words, each agent presents the solution, the algorithm is navigate by properly adapting the Gravitational mass and inertia mass. By the time, we will expect that masses be allure by the heavier mass. The mass will give Optimal solution in the space of the search.

When considering the singular system of masses, the communication between them is done by more certainly, masses obey the laws which are Law of Gravity and Law of motion stated [26]:

$$Fg = G(m_y \times m_z)/r2 \tag{3}$$

That is the gravitation law in between the two masses, m_y and m_z at some distance of r.

Here G is the function of time which reduces as increase of time as given in the equation:

$$G(t) = G_0(t) \times (t_0/t)^{\alpha}, \alpha < 1 \tag{4}$$

As the newton second law of motion,

$$acce = Fo/ma \tag{5}$$

Gravitational Search Algorithm is done on the basic laws. Rashedi [26] given according to the results of the experiments, given better results which are inversely proportional to distance when considered to square inverse proportionality given in the law of Gravitation. Hence the equation 3 is modified suitably. The gravity force causes all of the particles to move to other particles which are

with the heavier masses. Since of movement of masses occur globally, this is step which is explored. The heavier masses movement is explained because they are closer to the optimal solution and compared to the mass which are light slowly movable particles.

The position for the particle n is Pi

$$P_i = (p_e^1, p_e^2, \ldots, p_e^f, \ldots, p_i^n) \tag{6}$$

Here, $e = 1, 2, 3, 4 \ldots, n$ and p_e^f = position of particle $'e'$ in dimension $'f'$. The acting force from j^{th} mass on the z^{th} mass, at an time t is :

$$F_{ed}^f = G(t) \times ((M_e(t) \times M_d(t))/(R_{ed}(t) + \epsilon)) \times (p_d^f(t) - p_e^f(t)) \tag{7}$$

Based on the above force, The acceleration value(acce), Velocity(u) and position (p) of the agents are calculated as shown

Acceleration,

$$acce_e^f(t) = F_e^f(t)/M_e(t) \tag{8}$$

Velocity (u),

$$u_e^f(t+1) = r_j \times u_e^f(t) + acce_e^f(t) \tag{9}$$

Position (p),

$$P_e^f(t+1) = p_e^f(t) + u_e^f(t+1) \tag{10}$$

The randomly generated the number between 1 and 0 . In equation 7, $R_e^d(t)$ shows the particle distance. ϵ is the constant denominator should be added to of equation 7 to not show the error of the divide by zero.

The mass, m_e is computed as:

$$m_e(t) = (fitness_i(t) - worst(t))/(fitbest(t) - worst(t)) \tag{11}$$

$$M_e(t) = (m_e(t))/\sum_{d=1}^{N} m_d(t) \tag{12}$$

$$Fitbest(t) = max(fitness_d(t)), d\epsilon 1, 2, 3...N \tag{13}$$

$$worst(t) = min(fitness_e(t)), e\epsilon 1, 2, 3...N \tag{14}$$

The force acting on e particle in dimension f by the other particles is :

$$F_e^f(t) = \sum_{d\epsilon kbest, d \neq i}, rand_d F_e d^f(t) \tag{15}$$

In the above equation 15, the proper balance in between the exploitation and exploration, the k-best agents are taken for force applying on each another.k the time function reduces to 1. At the beginning all the particles apply force to each other. As the coefficients the weights of random are assigned.

By seeing the GSA interms of efficiency, points are noted [26].

- Here each mass can look at the efficiency of other mass, such that the GF is an Information transferring mechanism.
- The force from the nearest agents acts upon the mass, the agent can observe the area around it.
- The heavier mass has more attraction diameter, so the great effectiveness of attraction. Hence, the mass with more effective performance will have the most gravitational mass. As the above statement says so the agent will go to the better agent.
- The mass inertia is opposite to the motion and the movement of the mass slowdown. Hence, the agent moves slowly which have the inertia mass in heavy, and space search is done locally. So, adaptive learning is considered.
- The search accuracy is adjusted by the Gravitational constant.

Gravitational search algorithm is memory less but it does work like the algorithm with memory.

D. Binary GSA

There are many Evolutionary optimization problems such as feature selection and dimensionality reduction [27]–[30] in which is to do the solutions as binary vectors. In addition, problems in the dimensional space are taken in the binary space, too. The best solution is to show the digits whole numbers rather than in binary digits. The basic concepts of GSA are certainly modified in the Binary GSA. Here in the binary environment, every dimension can take only 0 or 1. Moving through the dimension means going through the 0 or 1. Mainly after the first iteration of the algorithm in the updation of the velocity here we use the

$$S(V_e^f(t)) = |tanh(V_e^f(t))| \tag{16}$$

$$f(x) = \begin{cases} p_e^f(t+1) = \ (1 - p_e^f(t)), & r_j < S(v_e^f(t+1)) \\ p_e^f(t+1) = p_e^f(t), & else \end{cases}$$

To achieve the binary version for the next iteration. The main observation between the GSA and BGSA is that the position updating is done using the binary version 0 or 1 values. It is done according to the velocity of mass and considering the value if the position which is less than the random value should be taken as 0 or if it is more than the random value it should be considered as 1. A small value of the velocity and position must provide the small probability of position changing. A large value of velocity should give large probability of the change in the mass position from its previous position. In other words, the least value of the velocity will provide the good mass position which should not be changed (Considering the least value is 0) it is noted the value is calculated hamming distance [31].

III. FS With Different optimization problems

A. Feature selection with Genetic Algorithm

FS in view of many problems in practical(e.g medical data) gives a part of the multi-situation efficiency criteria. The multi criteria should be efficiently done including the perfectness of the classification, the risk with classification which depends on the pattern description by attribute selection. EA gives a peculiar approach to multi-situational optimization condition. The wrapper based is used for the multi situational approach for subset selection using GA with faster distance based neural learning algorithm. The queries reader refer to [11] for the more clarification. These are simple computational elements. The feature selection using the genetic algorithm for neural network pattern classifiers is given by Jihoon Yang where traditional neural algorithms have the problems. Back propagation is the algorithm which takes on error search or weight in the space done by user specified pattern. If the Network contains lesser number of neurons then it takes, The algorithm not perfect to find the classification functions leads to poor generalization. In both the case, It is difficult to take the subset and give the training patterns which are used.The results are based on using the genetic algorithms [32], [33] with some strategy knowns as the rank based selection strategy. The solutions are 10 fold cross validation or the task of the classification taken some values; The size of population: 50, The generation number: 20, The probability of crossover: 0.6, The probability mutation: 0.001, and The selection probability: 0.6.

$$Fitness(a) = accu(a) - (cost(a)/(accuracy+1)) + Cost_m \quad (17)$$

Here fitness(a) denotes the fitness of the subset taken by a, $accu$ is the accuracy of the test trained, Costm is the max bound This gives that $\forall <= fitness(a) <= (100+cost_m)$. It is to combine non-tivil solutions based on the knowledge of domain.

In recent years many feature selection algorithms have been proposed by using GA [34]–[36].

B. Feature selection with PSO

The gene profiles, give the cell state at molecular level, which had great time in diagnosis tool, when the number of genes are high, training sets has a short size of sample in cancer type [37]. These sets of data combined a challenge to classification types. A trustable method of selection for relevant genes for the classification of sample is needed for speeding the processing rate, and reduce the error of predictive rate and to remove the redundancy by large number of genes involved for investigation. [37] used IBPSO to implement the feature selection and the k-Nearest Neighbor $(k--NN)$ method does as a evaluator for the problems of classification using PSO. The results show that this method reduces the number of features required and efficiently feature selection simplification is done.

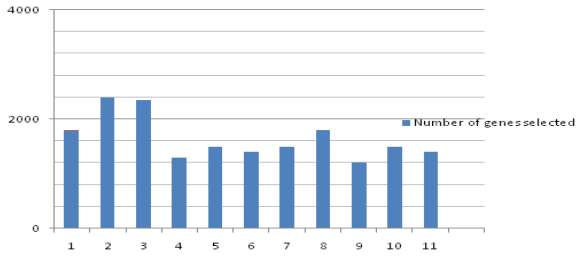

Fig. 2. Number of genes selected for 11 datasets

The accuracy of classification is obtained has the largest accuracy in available 9 of eleven gene data test problems, particle velocity are updated and tried to be maximum If the addition of the cause the dimension velocity to over maximum then dimension velocity is limited to some V_{max}, The V_{max} and V_{min} are user defined. The PSO rapidly converges through the starting stages of the search and slows down consequently and the particle get trapped in the optimum of local, the value of gbest is calculated before position particle is updated.If best value is same then particle will be trapped in optimum of local.

Fig. 2 describes Number of genes (features) selected for each of the 11 datasets. The data given includes the name of the dataset, sample numbers, categories, samples, genes selected, and gene selected percentage. The percentage of the average selected is 0.17 The lowest and the highest percentage of genes reduced to 0.12 and 0.24 for the data sets of tumours

C. Feature selection with GSA

The feature selection method based on GSA and OPF [38]. Here, main theme is to use the OPF accuracy as the fitness value over the set of evaluation to maximized Binary GSA. So, each and every agent is the probable solution in the dimensional space, where 1 value shows that feature is selected do the new data set and 0 value shows that feature is not selected. The FS algorithm is combines optimization of the GSA [26] with the OPF speed [38] classifier. Here [20] used the accuracy of the OPF as function to route GSA into searching the better solutions. The solution of the vector resulted by GSA is more trustable as the masses increases. So, there is need in classifier faster for training all possible subset of feature designated by each particle position. The algorithm taken donot stop the OPF as classifier, used for efficiency of training. It doesn't solve optimization problem in parameter, such as neural networks.Algorithm [20] is used to feature selection. The robustness assessment is done.

For the Evaluation of the data of medical for prediction of disease requires the techniques of feature selection which are efficient, the data have the huge number of the features. Research have done using EC(Evolutionary computation) such as GA [14], PSO [15] for feature selection and found them faster than the normal techniques. So [39] used the almost a new technique in the field of medical called GSA

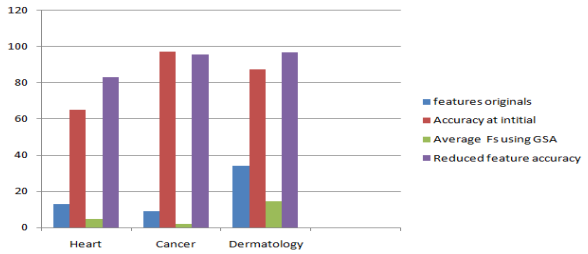

Fig. 3. Reduction rate

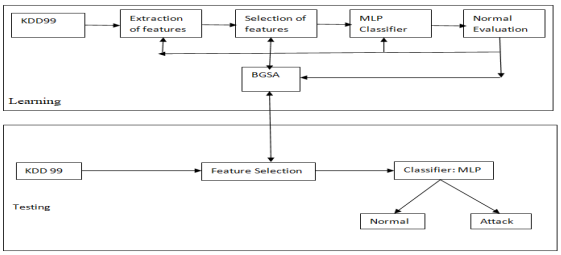

Fig. 4. Feature Subset Selection Using Binary Gravitational Search Algorithm with KDD99

for the selection of features in the datasets. The method of wrapper based is used, combining the GSA and K-nearest removes the unwanted data by average of 66% improving the prediction accuracy.

The fitness function is used for better position is :

$$Fitness = \alpha \times \gamma + \beta \times (T - S)/S \tag{18}$$

This experiments are done number of times to deduct the random variables.

The K-NN is used for the purpose of classification. The classification efficiency is measured on the accuracy,

$$accuracy = (TPO+TNE)/(TPO+TNE+FPO+FNE) \tag{19}$$

preventing the overfitting, the validation is done.

Fig. 3 showing the feature reduction is better i.e.., 64.61% and the accuracy has quietly increased from the range of 64.81% to 82.96% in the dataset of heart.The Dermatology set reduced to 57.64% and efficiency of classifier is 87.14% to 96.71%. In the breast cancer the higher value is obtained 77.77% and slight decrease of the accuracy 97.14% to 95.7%.

The feature selection is used with the GSA which is modified [40]. The modified GSA linear piecewise map of chaotic to increase the species diversity and the quadratic sequential accelerating of programming of local exploration. This work is to improvise and optimize the selection by the Modified GSA. Different experiments and comparisions [39] are done. The system has better performance and accuracy has higher which is achieved compared to available dataset and other systems.

The pseudo-code of the FSS-MGSA technique.

Algorithm 1 : The BPSO Algorithm for Feature Selection

Step: 1 Begin individual and position of a population
Step: 2 Do while For i = 1 to population size
Step: 3 Train classifier and evaluate fitness function
Step: 4 Update individual of the i^{th}object
Step: 5 change position of the i^{th}object
Step: 6 Update of the i^{th} object
Step: 7 Next i Until termination criterion is met End

This code is used to improvise the accuracy and the optimization in the datasets.

The Behjat A.R conducted the section of feature in the Security system as intrusion system. To control computers which are combines infrastructures this system has played a prominent role. To improvise the false rate at its lowest the different methods are used The BGSA [31] as the selection for features reduces the non-wanted features in KDD 99 detection of intrusion system and improvises the multilayer performance with lowest computation cost the increase has gone to nearly 100%.

The KDD system is used such that the intrusion and class 2 is used for better results. The attack and normal classes by NLP on the features selection method, settings are used. The size of population and iteration number are identified N=5 and T=20 respectively. The gravitational constant has been taken normal features are lowered the fitness function is used weight is defines such that most important role has been played in the feature selection size and the accuracy is improvised the intrusion has been detected and false response are reduced using the learning and testing sets such that vast improvement is done [41].

IV. Conclusion

In this paper presented a survey on feature selection problem solved by evolutionary optimization methods. Here, a new optimization method GSA is discussed in detailed and demonstrated feature selection with GSA. Some of the well known optimization methods have also discussed in view of feature selection problem. We compared these algorithms in feature selection problem.

References

[1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
[2] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, 2016.
[3] S. Das, "Filters, wrappers and a boosting-based hybrid for feature selection," in *Icml*, vol. 1, 2001, pp. 74–81.
[4] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
[5] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on knowledge and data engineering*, vol. 17, no. 4, pp. 491–502, 2005.

[6] S. Yu, S. De Backer, and P. Scheunders, "Genetic feature selection combined with composite fuzzy nearest neighbor classifiers for hyperspectral satellite imagery," *Pattern Recognition Letters*, vol. 23, no. 1-3, pp. 183–190, 2002.
[7] Z. Yan and C. Yuan, "Ant colony optimization for feature selection in face recognition," in *Biometric Authentication*. Springer, 2004, pp. 221–226.
[8] R. S. Wahono and N. S. Herman, "Genetic feature selection for software defect prediction," *Advanced Science Letters*, vol. 20, no. 1, pp. 239–244, 2014.
[9] H. M. Harb and A. S. Desuky, "Feature selection on classification of medical datasets based on particle swarm optimization," *International Journal of Computer Applications*, vol. 104, no. 5, 2014.
[10] H. A. Firpi and E. Goodman, "Swarmed feature selection," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*. IEEE, 2004, pp. 112–118.
[11] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature extraction, construction and selection*. Springer, 1998, pp. 117–136.
[12] Y. Rahmat-Samii and E. Michielssen, "Electromagnetic optimization by genetic algorithms," *Microwave Journal*, vol. 42, no. 11, pp. 232–232, 1999.
[13] J. Wilson, "Search methodologies: introductory tutorials in optimization and decision support techniques," 2007.
[14] K. Deb, "An introduction to genetic algorithms," *Sadhana*, vol. 24, no. 4-5, pp. 293–315, 1999.
[15] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*. Springer, 2011, pp. 760–766.
[16] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *ACM SIGGRAPH computer graphics*, vol. 21, no. 4. ACM, 1987, pp. 25–34.
[17] W. T. Reeves, "Particle systemsa technique for modeling a class of fuzzy objects," *ACM Transactions on Graphics (TOG)*, vol. 2, no. 2, pp. 91–108, 1983.
[18] M. M. Millonas, "Swarms, phase transitions, and collective intelligence," *arXiv preprint adap-org/9306002*, 1993.
[19] A. Kaveh, *Advances in metaheuristic algorithms for optimal design of structures*. Springer, 2016.
[20] S. Dara and H. Banka, "A binary pso feature selection algorithm for gene expression data," in *Advances in Communication and Computing Technologies (ICACACT), 2014 International Conference on*. IEEE, 2014, pp. 1–6.
[21] S. Dara and B. Haider, "An elitist binary pso algorithm for selecting features in high dimensional data," in *Advanced Computing, Networking and Informatics-Volume 1*. Springer, 2014, pp. 679–686.
[22] B. Haider and D. Suresh, "Hamming distance based binary pso for feature selection and classification from high dimensional gene expression data." in *IWBBIO*, 2014, pp. 507–514.
[23] H. Banka and S. Dara, "A hamming distance based binary particle swarm optimization (hdbpso) algorithm for high dimensional feature selection, classification and validation," *Pattern Recognition Letters*, vol. 52, pp. 94–100, 2015.
[24] S. D. Chandra Sekhara Rao Annavarapu and B. Haider, "Cancer microarray data feature selection using multi-objective binary particle swarm optimization algorithm," *EXCLI journal*, vol. 15, p. 460, 2016.
[25] B. H. Suresh Dara and A. C. S. Rao, "A rough based hybrid binary pso algorithm for flat feature selection and classification in gene expression data," *Annals of Data Science*, vol. 4, no. 3, pp. 341–360, 2017.
[26] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "Gsa: a gravitational search algorithm," *Information sciences*, vol. 179, no. 13, pp. 2232–2248, 2009.
[27] A. Pal and J. Maiti, "Development of a hybrid methodology for dimensionality reduction in mahalanobis–taguchi system using mahalanobis distance and binary particle swarm optimization," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1286–1293, 2010.
[28] M. Bereta and T. Burczyński, "Comparing binary and real-valued coding in hybrid immune algorithm for feature selection and classification of ecg signals," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 5, pp. 571–585, 2007.
[29] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary pso for feature selection using gene expression data," *Computational Biology and Chemistry*, vol. 32, no. 1, pp. 29–38, 2008.
[30] X.-P. Zeng, Y.-M. Li, and J. Qin, "A dynamic chain-like agent genetic algorithm for global numerical optimization and feature selection," *Neurocomputing*, vol. 72, no. 4-6, pp. 1214–1228, 2009.
[31] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "Bgsa: binary gravitational search algorithm," *Natural Computing*, vol. 9, no. 3, pp. 727–745, 2010.
[32] J. Holland and D. Goldberg, "Genetic algorithms in search, optimization and machine learning," *Massachusetts: Addison-Wesley*, 1989.
[33] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.
[34] H. Banka and S. Dara, "Feature selection and classification for gene expression data using evolutionary computation," in *Database and Expert Systems Applications (DEXA), 2012 23rd International Workshop on*. IEEE, 2012, pp. 185–189.
[35] M. Elloumi and A. Y. Zomaya, *Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data*. John Wiley & Sons, 2013, vol. 23.
[36] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 309–313, 2015.
[37] C.-S. Yang, L.-Y. Chuang, Y.-J. Chen, and C.-H. Yang, "Feature selection using memetic algorithms," in *Convergence and Hybrid Information Technology, 2008. ICCIT'08. Third International Conference on*, vol. 1. IEEE, 2008, pp. 416–423.
[38] J. P. Papa, A. X. Falcao, and C. T. Suzuki, "Supervised pattern classification based on optimum-path forest," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.
[39] S. Nagpal, S. Arora, S. Dey *et al.*, "Feature selection using gravitational search algorithm for biomedical data," *Procedia Computer Science*, vol. 115, pp. 258–265, 2017.
[40] X. Han, X. Chang, L. Quan, X. Xiong, J. Li, Z. Zhang, and Y. Liu, "Feature subset selection by gravitational search algorithm optimization," *Information Sciences*, vol. 281, pp. 128–146, 2014.
[41] A. R. Behjat, A. Mustapha, H. Nezamabadi-pour, M. N. Sulaiman, and N. Mustapha, "Feature subset selection using binary gravitational search algorithm for intrusion detection system," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2013, pp. 377–386.