Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 118 (2018) 50661-50665

ISSN: 2229-712X

Impact Meta Heuristic Algorithm on Collection Data Machines Learning to Data Mining Methods

MaysamToghraee[1], Marziyeh Bahrami[2], Mohammad esmaeili[2] and Farhad Rad[3]

[1]Department, Software Computer Engineering, Science and Research, Islamic Azad University, Yasouj, Iran.
[2]Department, Software Computer Engineering,Science and Research, Islamic Azad University, Ardebil, Iran.
[3]Faculty of Engineering, Department of Computer Science, Islamic Azad University, Yasouj, Iran.

ARTICLE INFO

Article history:
Received: 06 January 2018;
Received in revised form:
1 May 2018;
Accepted: 11 May 2018;

Keywords
Feature selection,
Data mining,
Algorithm cluster,
Heuristic methods,
NWINE data.

ABSTRACT

Intrusion detection system shave been around for quite some time, to protect systems from inside and outside threats. Researchers and scientists are concerned on how to enhance the intrusion detection performance, to be able to deal with real-time attacks and detect them fast from quick response. One way to improve performance is to use minimal number of features to define a model in a way that it can be used to accurately discriminate normal from anomalous behaviour. Many feature selection techniques are out there to reduce feature sets or extract new features out of them. In this paper, we propose an anomaly detectors generation approach using Meta heuristic algorithm in conjunction with several features selection techniques, including principle components analysis, sequential floating, and correlation-based feature selection. Meta heuristic algorithm was applied with deterministic crowding niching technique, to generate a set of detectors from a single run. In this test, based on various algorithms, we conclude that **NWINE** data is low in accuracy and only in the clustering algorithm, which rises precisely.

1. Introduction

Vast dimensions management is one of the common challenges for extracting knowledge and machine learning. Feature selection methods is one of the most challenging and the most important activities in developing machine learning and patterns recognition. Feature selection is one of the issues which has been discussed in machine learning and also pattern statistical recognition. This issue is important in many usages(e.g. classification), since there are a lot of features in these usages. Many of them are useless or are not informative**.**

Eliminating these features does not change the informative content but it Effects on calculating feature of the mentioned usage Also it helps to save much useless information with useful data. Not elimination of the waste features makes some dimensional problems. Dimensional problem says that when the dimensions increase, it is possible that the 2 data (or sample) get far from each other. Distance between those samples are estimated much more.

It makes the distance between both samples less representing the real distances. So, the quality of classifying or clustering are unpleasantly unreal and drop. It can be stated in another way. It can be said that some clusters or branches in feature's atmosphere are more coherent with some special features; Three general ways have been submit to overcome the above dimension problem: (a)using subspaces determined for clusters or branches by user, (b) using feature selection methods or decreasing dimensions like analyzing main factors and finally (c) using subspace clustering or subspace classifying methods. We discuss about the feature selection methods (b) in this report. A lot of solutions and algorithm have been represented for feature selection issue.

A lot of solutions and algorithms have been represented for feature selection issue, some of which are 30 or 40 years old. The problem about algorithms when represented was their calculating feature. However, fast computers and big saving sources have made this problem unimportant, beside, big data sets for new issues has made it important to find a fast algorithm for this issue. Feature selection has 2 types :

a) **supervised feature:** Labels are used during feature selection algorithm(Zhao &Liu 2007)

b) **unsupervised feature:** Labels are not used during feature selection algorithm(G.D 2008)

Research domain is just limited to the supervised feature selection while labels are used during feature selection algorithm .

2. Efficience Function

To calculate the efficiency function, we should first calculate the relationship of each feature with other features and label.

After calculation of each feature with other features and label, selecting features is done regarding to relationship amount of each feature with other features and labels. It means that the higher the dimensions(features), probably the most distance between them randomly. As the result, those samples are affected by the dimension and the has the most relationship with label and second the selecting features have the least relationship with other selecting features. Both are demonstrated in evolutional algorithms of efficiency function explicitly . Now, we use the following relation which shows the variation of selecting features and label similarity of selecting features to calculate the efficiency of this chromosome.

Tele:
E-mail address: may.toghraee@yahoo.com

$fit_{ch} = \sum_{i=1}^{f} lesser\left(max_{k=1}^{f}\left(|cor(X_i, X_k)| \times and(ch_i, ch_k)\right), th_2\right) \times \alpha + greater\left(|cor(X_i, T)|, th_1\right) \quad (1-2)$

Where f is the amount of features and α is the big positive number, th_1 and th_2 are two thresholds which should be adjusted by the user, ch_j shows *I*th of chromosome , | | shows the absolute value, $cor(X_j,X_k)$ shows the relationship of I th and k th features , and (ch_j,ch_k) show the logical operator (output is 1 when both inputs are 1,otherwise function output is 0) , *(a,b)* is greater than1 if $a\geq b$,otherwise function output is 0 and function*(a,b)* is less than 1, if $a\leq b$. otherwise function output is 0(toghraee .M& ET al; 2016) .

3. Evaluation Algorithms

3.1 Genetics Algorithm

The main idea of evolutional algorithms were presented in 1960 by Ritchenberg. Genetic algorithms are derived from this type of algorithm. In fact the computer search methods based on optimization algorithms based on the genetic structure of chromosome's by john Holland(1970) was introduced at the university of Michigan(A.mehdi;1386). The most extensive definitions of genetic algorithm are from goldberg's book:" Genetic algorithms is machine learning model, its behavior is an example of the evolutionary processes in nature". Genetic algorithm is one the strongest methods derived from the nature which seeks for the problem A searching bee chooses the food source regarding the possibility related to that source, P_i , which is calculated by the following phrase:

based on (C . Meyers &et al; 1996): First we answer the question by defining a chromosome structure(coding). Introducing the fitness function, we explain the quality of the given answers in each chromosome numerically. Then we generate some chromosomes randomly(or semi-random. And the chromosomes are known as the initial population. We have some answers for the problem in this step which have lower quality. The quality of each chromosome of the population is specified according to fitness function we specified. Now we select two chromosomes for reproduction using an appropriate method (A method in which the probability of chromosome selection with better fitness amount is more than a chromosome with less fitness amount). Then using these two chromosome, we create a new chromosome(mating). We change some genes of some chromosomes having a specified probability. Selecting, recombination and mutation steps make a new population of chromosomes(new generation). If the chromosomes tend to the demanded answer ,reproduction process stops. Otherwise creating a generation out of the previous generation continues until we reach to a desired answer or ending the algorithm, condition .

3.2 Particles Swarm Optimization Algorithm

In 1995, Kennedy and Eberhart offered particle swarm optimization algorithm for the first time as a non-deterministic search methods for functional optimization (Kenndy &et al,1995). This algorithm has got inspired from the collective movement of the birds that were seeking for food. A group of birds are looking for food in a space randomly. There is only one piece of food in the discussed space. None of the birds know the location of food. One of the best strategies can be following a bird who is closer to the food. This strategy is the basis of the algorithm (M. carvalho& et al, 2006). Each solution, which is called a particle, is the same as a bird in the algorithm of a mass movement of birds. Each particle has a fitness value which is determined by a fitness function. Particle swarm works based on this principle. In each moment each particle sets its location according to the best location in which it has been located in the searching atmosphere and in the best location which is in its neighborhood. To do this, methods depended on evolutional algorithms have been represented for selecting subsets of features in this chapter we discuss about the efficiency function of this algorithms. We are seeking for 2 targets in feature selecting . First, an evolutional algorithm is looking for subset of features which position of each particle, a new position can be considered for the particle. The function of updating the position of particle is as below :

$$X_i^{(t+1)} = X_i^{(t)} + V_i^{(t+1)}, \text{where } X_i^{(t)} \sim u(X_{min}, X_{max})$$

(3-2-1)

The suitability of a particle in the search space is evaluated by the fitness function .

3.3. Colony Bees Artificial Algorithm

Ants algorithms based on intellectual foundation can be simply stated in one sentence: Ants select the best way between different ways of reaching food among the barriers in nature. The short way is always chosen. Ants secrete a substance called pheromone after finding food which is seen white after the rain. They find their way finding the pheromone way. Bees are in three groups in colony bees algorithm bees:

Worker bees, the audience and the pioneers(scout). A honey bee stays in dance region to make a decision for choosing a food source is called searching honey bee and and a honey bee, which is looking for the specified is called worker bee. A bee which searches randomly is called pioneer honey bee(scout). In the bees algorithm, worker bees are half of the population and the other half are searching bees. For every food source there is only one worker bees, in other words, the number of worker bees around the hive equals the number of food sources. The worker bees who are tired of working in the food supply are leading wasps browser. The main steps of the algorithm is given below:

- **Initialization**
- **Repeat**

a) The location of worker bees in food supplies in memory
b) The location of searching bees in food sources in memory
c) Sending the pioneer wasps bees to search for new food sources.

- **(Until the desired situation gets achieved).**

3.4. Big Bang Algorithm

First algorithm was introduced by Eksin and Erol (I.Eksin&K.Erol,2006). This algorithm uses the phenomenon of the big bang and then the contraction of the universe in the center of gravity. This algorithm has higher speed of convergence compared to other algorithms. Particles are scattered in searching space randomly like other evolutional algorithms which is called big bang phase (c.v. camp,2007). each particle has a position which determines what is the particle coordinates in the search space. Position of the particle changes by its movement in the time passing. $X_i^{(t)}$ determines the ith position of the particle in the t^{th} time. Also, each particle needs a speed for moving in the space.

$V_i^{(t)}$ is the speed of i^{th} particle in t^{th} time. Adding speed to the phase, all the particles accumulate around the center of gravity. This phase acts like a converging operator which is calculated through the following equation :

$$X_i^{(k)} = \frac{\sum_{j=1}^{N} \frac{x_i^{(k,j)}}{f_j}}{\sum_{j=1}^{N} \frac{1}{f_j}} \qquad i = 1,2,3,4,\dots,c$$

(3-4-1)

Where $X_i^{(K)}$, is the i[th] particle of the center of gravity in k^{th} repetition. $X_i^{(k,j)}$ is the I[th] component of j'[th] particle produced in k'[th] repetition. f_j is the target function for point j and N is the number of points or particles and C the number of controlling variables respectively, after determining the particles center of gravity, the new position of particles can be calculated using the following equation :

$$X_i^{(k+1,j)} = X_i^{(k)} + (X_{imax} - X_{imin}) \times \alpha_1 \times r \times \left(\frac{1}{k+1}\right)$$

(3-4-2)

r is the random number, X_{imax} and X_{imin} respectively are lower and upper constraints to limit in the above equation . α_1 is also a parameter for limiting the search space.

3.5. Hill climbing Algorithm

In Hill climbing algorithm, at first an answer to the question is generated randomly. Then in a loop until the stop condition is not established for the algorithms, a number of neighbors are generated for the current mode of production and the best one is chosen among them and replaces the current mode. (of course, another definition for the hill climbing has been stated). In general, the optimality of the answer to the algorithm is local. To run Hill climbing algorithm we need two functions : performance function and neighbor function. The performance function determines optimality of the answer. 8 minister of guards on board pairs queens returns. Neighbor function also produces current mode neighbors. In problem 8 queens are for generating neighbor modes, each of them are chosen and move once upward or downward . This means that in the worst mode of each case, there will be 16 neighboring modes that in each repetition of the loop as the best answer will be replaced neighbors. ends when there is not a better mode compared to the current mode.

Ants algorithms based on intellectual foundation can be simply stated in one sentence: Ants select the best way between different ways of reaching food among the

3.6. Clustering algorithm

A general trend for clustering process includes the following steps:

1. Displaying the patterns that usually involves selecting or extracting the feature.
2. Defining an assessment criteria of the similarity according to the data domain.
3. Clustering or Grouping process.
4. Summarization of data if needed.
5. Validation of system.

Displaying the patterns refers to the number of classes, available samples and the number, type and scale of features in a clustering algorithm. Some of these data are not controlled by the user. Selecting feature, Identifying process of a subset is one of the most effective features for using in clustering and feature extracting is, the process of changing some available features and generating new features. Both of these techniques are used in order to achieve a suitable set of features and enhancing the performance of clustering. Adjacent of samples usually is measured by a function of distance between the pair of input patterns.

Various criterion are used to measure distances in various fields (toghraee M. & ET al ; 2017).

4. Evaluation Methods

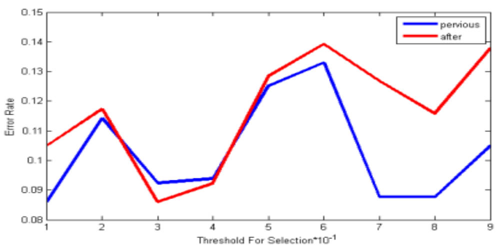
In this section, the results of applying the proposed method on different data sets and used parameters has been reported. In this section, analysis and interpretation of the results and efficiency of the proposed algorithm also briefly discussed. In this research validation includes experimental validation. In experimental validation, the efficiency of proposed algorithm is discussed in comparison with other algorithm in several real data sets. UCI standard data sets are the used datasets(it is on the machine learning website) is that almost all the result of recent studies in the field of data mining in the world are reported using this data set (Toghraee .M & Et al; 2016). These actual data are often standard. Experimental results of the proposed method and other methods with the world's valid criterion such as error size the feature set size and chosen will be reported.

Table1. datasets used in the first experiment in this thesis. Starred data set.

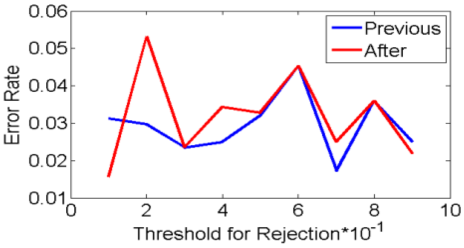
Dataset Name	*# of data items*	*# of features*	*# of classes*	*Data distribution per clusters*
*Breast Cancer**	*404*	*9*	*2*	*444-239*
*Bupa**	*345*	*6*	*2*	*145-200*
*Glass**	*214*	*9*	*6*	*70-76-17-13-9-29*
*Galaxy**	*323*	*4*	*7*	*51-28-46-38-80-45-35*
*SAHeart**	*462*	*9*	*2*	*160-302*
*Ionosphere**	*351*	*34*	*2*	*126-225*
Iris*	***150***	***4***	***3***	***50-50-50***
NWine*	***178***	***13***	***3***	***59-71-48***
*Yeast**	*1484*	*8*	*10*	*463-5-35-44-51-163-244-429-20-30*

It should be mentioned that all results represented in this thesis are achieved, because the test results are so strong and extendable, with an average of 10 separate performance. In the first experiments have been performance on several real data sets. The actual data sets have been derived from machine learning website (Newman,1998). These data sets is presented in the table above. We tried the data sets have variety in the number of classes, features and samples in doing experiments, until the test results be at the maximum amount of strength and be extendable. Test results of the standard features of these data sets have been reported. in other words, each of the properties of this dataset with zero mean and variance one, N(0,1) are standard. Thresholds th_1 and th_2 changing from 0,1 to 0,9 and training set size from 10 percent to 90 percent, different levels of accuracy are achieved. Notice that each performance is done10 times repeatedly. In all cases the maximum population size should be adjusted on 50 and the maximum generation size on 50(Toghraee .M., et al ;2017) .

5. Results from Algorithms



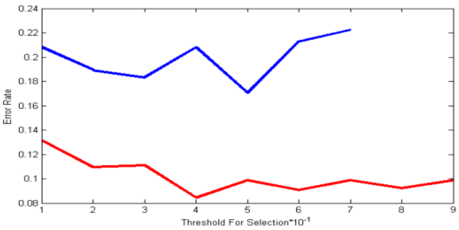
a)Selecting the threshold of error rate.



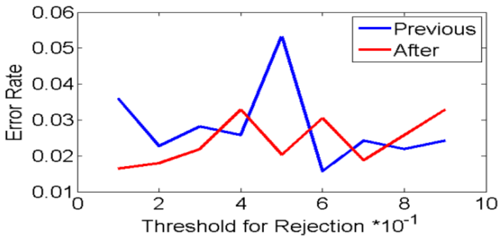
b)Rejecting the threshold error rate.

Fig 1. Selecting and rejecting threshold error rate on the data set Nwine in Genetic algorithm.

In genetic algorithm on dataset Nwine , in selection mode, pervious error and after error rate situation in this data than to other data is better. Because they are nearly. In threshold 0.4 situation error rate almost is 0.09 and in threshold 0.7 to 0.8 pervious error rate is low. But in mode rejection, threshold 0.6 and 0.8 have collision but error rate have high.



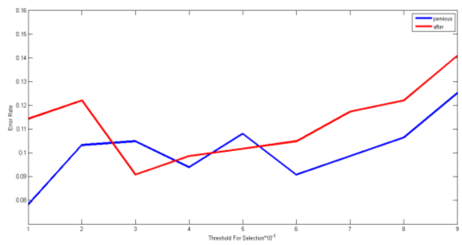
a)Selecting the threshold of error rate.



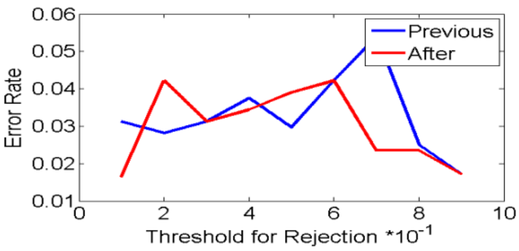
b)Rejecting error rate threshold.

Fig 2. Selecting and rejecting threshold error rate on the data set Nwine in Hill climbing.

In hill climbing algorithm on the data set Nwine, situation pervious error and after error isn't favorable. And pervious error rate and after error rate from are ware. In mode (b) pervious error rate and after error rate are nearly and situation than to mode (a) is better. Because in threshold 0.6 and 0.4 pervious error rate is low.



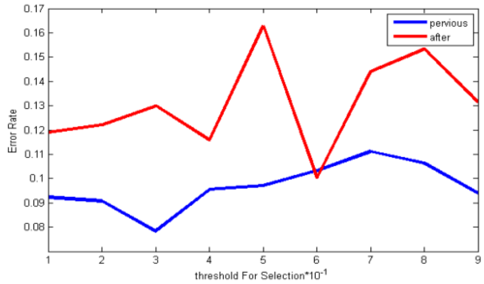
a)Selecting the error rate threshold.



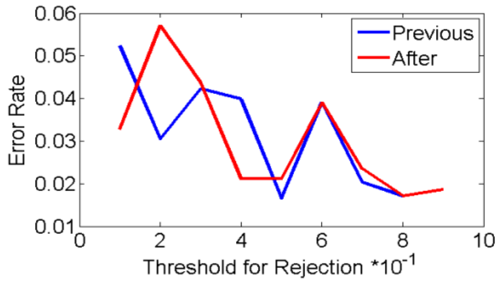
b)Rejecting error rate threshold.

Fig 3. Select and reject threshold error rate on the data set NwineArtificial bee.

In bees algorithm on the data set Nwine, situation in mode (a), pervious error rate and after error rate is favorable and they have nearly, and in threshold 0.1 minimum has pervious error rate . But in mode(b) error rate just in threshold 0.8 to 0.9 have collision is almost 0.018, and situation just in this point is better.



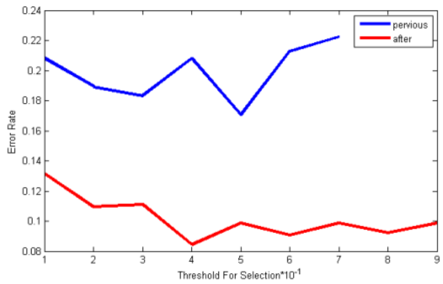
a) Selecting the error rate threshold.



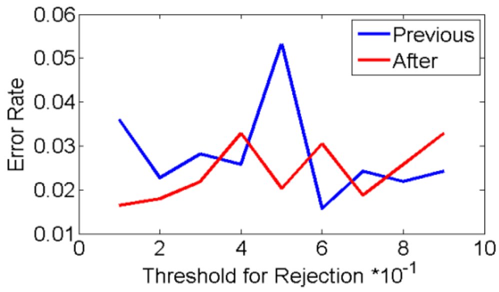
b)Rejecting error rate threshold.

Fig 4. Selecting and rejecting error rate threshold on the data set Nwine in Big Bang algorithm.

In the big bang algorithm on dataset Nwine in mode (a), previous error and after error just in threshold 0.06 has collision and in other threshold isn't better. But in mode(b) situation pervious error rate and after error rate from threshold 0.5 to 0.8 are nearly. But error rate is high. Just in threshold 0.8 situation were better.



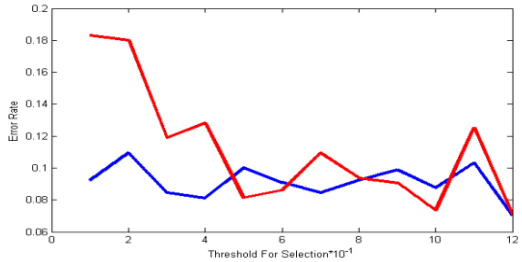
a)Selecting the error rate threshold.



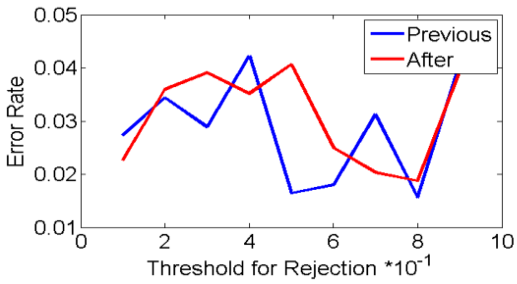
b)Rejecting error rate threshold.

Fig 5. Selecting and rejecting error rate threshold on data set Nwine in particle swarm optimization algorithm.

In particle swarm optimization algorithm on the data set Nwine, in state select, situation isn't favorable. But reject mode, pervious error rate and after than to made (a) is better, And in threshold 0.6 situation pervious error rate is very better.



a)Selecting the error rate threshold.



b)Rejecting error rate threshold.

Fig 6. Selecting and rejecting error rate threshold on the data set Nwine in cluster algorithm.

In cluster algorithm on data set Nwine, in mode (a) pervious error and after error situation in threshold 0.1to 1.2 error rate have low, but in this situation on the data set Nwine in this cluster algorithm then other algorithms are better. Because in this mode, errors are nearly. after error rate and pervious error rate in threshold 1.2 situation were very better. In mode (b), in threshold 0.8 situation errors are very low and situation is better. From these experiments obtained, we conclude that the situation clustering algorithm is better.

6. Conclusion

In this study, the efficiency of the proposed algorithm , advantages, challenges and innovations have been discussed briefly. A set of feature selecting method based on the collective intelligence methods has been presented. It was shown experimentally that these methods can have no decrease in classifying quality ,selecting almost 80% of features .Beside ,these experiments showed that clustering method is the best way in finding appropriate features for classifying .

7. Future Works

As future works , we should accomplish the data sets on hierarchical and partition clustering algorithm as it selects the final cluster qualitatively for us.

Reffrence

1.Camp C.V. "*Design of space trusses using Big Bang – Big Crunch optimization*", Journal of Structural Engineering, vol.133, Issue 7, (2007). pp.999-1008.

2.Dy J. Unsupervised feature selection. *Computational Methods of Feature Selection*, (2008). pages 19-39.

3.Erol Osman k and Eksin I. *"New optimization method : Big Bang-Big Crunch*", Elsevier, Advances in Engineering Software .(2006).37 pp: 106–111.

4.Guyon I and Elise A. *An introduction to variable and feature selection*. Journal of Machine Learning Research, (2003).3:1157-1182.

5.Jain k ,. Dubes R. C. *Algorithms for Clustering Data, Prentice Hall*, Englewood Cliffs. (2012).

6.Karaboga D, Ozturk C. A novel clustering approach: *Artificial Bee Colony (ABC) algorithm*. (2011).

7.Kennedy J and Eberhart R, C. *"Particle Swarm Optimization"*, Proceedings of IEEE International .

8.Lampinen J and Laksone J and Oja E. *Pattern recognition*. In editor, Image Processing and Pattern Recognition, volume 5 of Neural Network Systems .(2014).

9.Ladha L and Scholar R, Depa T, L*Feature Selection Methods and Algorithms*, International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 5, .(2011). pp. 1787–1797.

10.Marki F and Vogel M and Fischer M. *"Process Plan optimization using a Genetic Algorithm*", PATAT, (2006), pp. 528–531. ISBN 80-210-3726-1.

11.Marki F and Vogel M and Fischer M. *"Process Plan optimization using a Genetic Algorithm*", PATAT, (2006). pp. 528–531. ISBN 80-210-3726-1.

12.Perzina R. *"Solving the University Timetabling Problem with Optimized Enrolment of Students by aParallel Self-adaptive Genetic Algorithm*",(2006*)*. PATAT 2006, pp. 264–280. ISBN 80-210-3726-1.

13.Pham D. T and Ghanbarzadeh A and Koc E nad Otri S and Rahim S., Zaidi M. *The Bees Algorithm –A Novel Tool for Complex Optimization Problems*. (2006).

14.Susana M and Vieira J and Sousa M.C. *Fuzzy criteria for feature selection*, Fuzzy Sets and Systems, (2012).Vol. 189, No. 1, pp. 1–18.

15.Toghraee M, rad F, parvin H,.(2016*). THE impact of feature selection on meta heuristic algorithm to data mining methods.* International journal of modern education and computer science. Volum 8; issue 10., page(33).

16.Toghraee M, rad F, parvin H,.(2016). *Evaluation of meta heuristic algorithm for stable feature selection*. I.J.iformation technology and computer science (ijitcs),.volum8. issue:2074-9015; pp: 22-29.

17.Toghraee M, rad F, parvin H,.(2016). *Effect neural networks on selected feature by meta heuristic. i.j. mathematical science and computing (ijmcs).*volum2.issue: 2310- 9033.pp:41-48.

18.Toghraee M, Esmaeili M , parvin H,.(2016*).evaluation neural networks on selected feature by meta heuristic algorithms*. Artificial intelligent system and machine learning.volum8.pp:108-115.

19.Toghraee M, rad F, parvin H,.(2017). *Evaluation average total data set learning machine on the meta heuristic algorithm.* International journal of emerging trend& technology in computer science .volum6.issue:2278-6856.page(7).

20 .Toghraee M, rad F, parvin H,.(2017). *The Influence Select Feature on The Clustering Algorithm*. Journal of Computer Science Engineering and Software Testing Volume 3 Issue 3.pp:1-10.

21.Zhao Z and Liu H. *Semi-supervised feature selection via spectral analysis.* In Proceedings of SIAM International Conference on Data Mining (SDM). (2007).