EPH - International Journal of Science And Engineering

ISSN (Online): 2454 - 2016
Volume 06 Issue 03 September 2020

DOI: https://doi.org/10.53555/eijse.v6i3.73

A RELATIVE INVESTIGATORY ANALYSIS ON ANT COLONY ALGORITHM AND GENETIC ALGORITHM FOR FEATURE SELECTION

Udit Narayan Kar[1]*, Arunima Kar[2], Anyatama Kar[3]
*[1]*Computer Science Department, Saurashtra University, India*
[2,3]*Computer Science Department, Institute of Technical education and Research, India*

Corresponding Author:-

Abstract: -

Feature subset selection is used as a common technique in data pre-processing for pattern recognition, machine learning and data mining has attracted much attention in recent years. A good feature selection method can reduce the cost of feature measurement and increase classifier efficiency and classification accuracy. One approach in the feature selection area is employing population based optimization algorithms such as Genetic Algorithm (GA)-based method and Ant Colony Optimization (ACO) based method. In this paper ant colony optimization algorithm (ACO) is compared to Genetic algorithm (GA) using feature selection. Finally a comparative study is done to know the pros and cons of the work done.

Keywords: - *GA, ACO, Selection, Crossover, Mutation, Elitism*

1. INTRODUCTION

For many pattern classification problems a higher number of features used do not necessarily translate into higher classification accuracy. In some cases the performance of algorithms devoted to speed and predictive accuracy of the data characterization can even decrease. The presence of unimportant and superfluous features in datasets motivates researchers to devise novel feature selection strategies [6]. The purpose of the feature selection is to reduce the maximum number of irrelevant features while maintaining acceptable classification accuracy. A good feature selection method can reduce the cost of feature measurement, and increase classifier efficiency and classification accuracy. Feature selection is of considerable importance in pattern classification, data analysis, multimedia information retrieval, medical data processing, machine learning, and data mining applications. Due to the development of information acquirement and storage, tens, hundreds, or even thousands of features are acquired and stored in databases for some real-world applications. With a limited amount of training data, an excessive amount of features may cause a significant slowdown in the learning process, and may increase the risk of the learned classifier to over-fit the training data, because irrelevant or redundant features confuse learning algorithms [3],[5].

One approach in the feature selection area [7, 8] is employing population-based optimization algorithms such as Genetic Algorithm (GA)-based method and Ant Colony Optimization (ACO) based method [19]. ACO algorithm is inspired of ant's social behavior in their search for the shortest paths to food sources. Most common techniques for ACO-Based feature selection use the priori information of features. Feature selection methods seek to remove the redundant features, reduce the dimensionality of the feature space, and improve the accuracy of learning [1, 2]. In general, feature selection can be seen as an optimization problem in the feature space. Many Search algorithms have been proposed to reduce the computational complexity such as branch and bound SFS SBS Floating and GA. The GA is a powerful feature selection tool especially when the dimensions of the original feature set are large. Reducing the dimensions of the feature space not only reduces the computational complexity, but also increases estimated performance of the classifiers. The use of genetic algorithms search techniques for feature selection is not new. Many researchers have used GAs in feature selection. Finding the optimal feature selection is an NP hard optimization problem that involves searching the space of possible feature subsets to identify the optimal one. There are 2^n states in the search space (n is the number of features in the dataset). For large values, evaluating all the states is computationally infeasible. Therefore, many heuristics such as genetic algorithms (GAs) tab search (TS), simulated annealing (SA) and ant colony optimization algorithms (ACO) have been used for solving feature selection. The role of feature selection is to reduce the dimensionality of a feature vector by removing useless, redundant, or least useful features [11, 9]. After analyzing both GA and ACO algorithms and doing a fair bit of literature survey the authors found that in the existing algorithms it is difficult to maximize the overall information gain associated with a feature subset and the mutual co-relation between the among the feature is not fully minimized. In this paper the authors have given a proposed model and depending upon the proposed model and certain datasets the performance of GA and ACO are compared.

The rest of the paper consists of section 2 defining the fundamentals of both population of feature selection methods consisting of both GA and ACO, section 3 defines the problem statement, in section 4 a proposed model is given for both GA and ACO, in section 5 complete experimental results are analysed and paper ends with a fair conclusion in section 6.

2. Population based Feature Selection

Feature selection is a process that selects a subset of original features. The optimality of a feature subset is measured by an evaluation criterion. As the dimensionality of a domain expands the number of features N increases. Finding an optimal feature subset is usually intractable and many problems related to feature selection have been shown to be NP-hard. The GA is a powerful feature selection tool especially when the dimensions of the original feature set are large. Reducing the dimensions of the feature space not only reduces the computational complexity, but also increases estimated performance of the classifiers. The use of genetic algorithms search techniques for feature selection is not new. Many researchers have used GAs in feature selection.

2.1 .GA BASED FEATURE SELECTION

GA is a stochastic optimization procedure. It maintains a set of solutions called a population, to be analyzed and improved through successive iterations called generations. Each solution a chromosome is a string represented by a finite sequence of 0's and l's. GAs search and optimization algorithms based on natural evolution and selection as a means of determining an optimal solution to feature selection problem. GA starts with an initial random population; each individual in the population represents a candidate solution to the feature subset selection problem [20]. Each solution is of a predefined length of binary vector. If a bit is a 1, it means that corresponding feature is selected, 0 otherwise. the evolutionary process of GA composes of many biological limitations, such as the chromosome representation, genetic operators, population selection and fitness function. The main advantages of the GA are summarized as: simple concept, easy implementation, parallelism, and computational efficiency when compared with other mathematical algorithms. GA has been successfully applied to many complex optimization problems, and the efficiency of the algorithm has been confirmed.

2.1.1 Selection (or Reproduction)

The selection operator involves randomly choosing members of the population to enter a mating pool. The operator is carefully formulated to ensure that better members of the population (with higher fitness) have a greater probability of being selected for mating, but that worse members of the population still have a small probability of being selected. Having some probability of choosing worse members is important to ensure that the search process is global and does not simply converge to the nearest local optimum. Selection is one of the important aspects of the GA process, and there are several ways for the selection: some of these are Tournament selection, ranking selection, and Proportional selection. In the proportional selection a string is selected for the mating with a probability proportional to its fitness. There are many ways of proportional selection: the most popular are Roulette Wheel Selection (RWS), Stochastic Reminder Roulette Wheel Selection (SRRWS), and Stochastic Universal Sampling (SUS).

2.1.2 Crossover

Crossover creates a new individual's representation from parts of its parent's presentations. During crossover, pairs of chromosomes (parents) are randomly selected from the mating population. With a user-specified crossover probability Pc, genes from one parent chromosome are swapped with corresponding genes on the other parent chromosome to create two children. When the swap does not occur (probability 1- Pc), the two parents are transferred to the child population unchanged. In multipoint crossover multiple locations on the chromosome are selected for gene exchange, each with probability Pc. The highest amount of ex- change occurs during uniform crossover, where every gene has a probability Pc of being exchanged with its corresponding gene on the other parent chromosome.

2.1.3 Mutation

Mutation is a genetic operator used to maintain genetic diversity from one generation of a population of chromosomes to the next. It is analogous to biological mutation. Once the children are created during crossover the mutation operator is applied to each child. Each gene has a user-specified mutation probability, Pm, of being mutated. In binary mutation, a value of 0 converted to a value of 1, or vice versa.

2.1.4 Elitism

Elitism involves replacing worst chromosomes in the children population with the best members of the parent population. This operator has proved to increase the speed of convergence of the GA because it ensures that the best solution found in each generation is retained. While this operator could be applied more broadly e.g. retaining the 2 or 3 best solutions

2.2 ANT COLONY OPTIMIZATION

Marco Dorigo along with his colleagues proposed Ant Colony Optimization in 1992, based on the behavior of ants depositing pheromone to seek the shortest path between their colony and source of food [19]. Individual ants are capable of doing simple works but not complex tasks such as finding the shortest path to food source or optimal nest structure .Therefore they work in groups and communicate with each other using through the environment (pheromone deposits). An Ant finds a food source, F at some distance from its nest, N and returns back by some path. The ant leaves a pheromone trail behind it while coming back. Initially ants move through all the possible four paths. But the intensity of the pheromone deposited is strong for shorter path than the longer paths. Eventually, most of the ants follow the shortest path route from nest to the food source and back and slowly, the other paths lose their pheromone trails. This unique behavior of ant is implemented in the ACO to solve the real world problems. The first application of Ant Colony Optimization algorithm was for Travelling Salesman Problem, to calculate the shortest path between the cities by visiting each city exactly once. The ACO has been inspired by the behavior of real ants. It was observed that real ants were able to select the shortest path between their nest and food resources, in the existence of alternate paths between the two. While travelling their way ants deposit a chemical substance called pheromone on the ground when they arrive at a decision point, they make a probabilistic choice biased by the intensity of pheromone they smell. When they back the probability of choosing the same path is higher (due to increase of pheromone).They new pheromone will be released on the chosen path. This behavior has an autocatalytic effect because the very fact of choosing a path will increase the amount of pheromone on the corresponding path which in turn will make it more attractive for future ants to follow. Shortly all ants will select the shortest path.

2.2.1 ACO Method

ACO method is the random search method. In general, an ACO method can be applied to any combination problem as for as it is possible to define an appropriate problem representation. The problem can be described as a graph with a set of nodes and edges between nodes. Heuristic desirability (P) of edges .A suitable heuristic measure of the goodness of paths from one node to every other connected node in the graph Construction of feasible solutions. A mechanism must be in place where by possible solutions are efficiently created and A suitable method of updating the pheromone levels on edges is required with a corresponding trail update Rule.

Figure-1: Ant behavior basic view

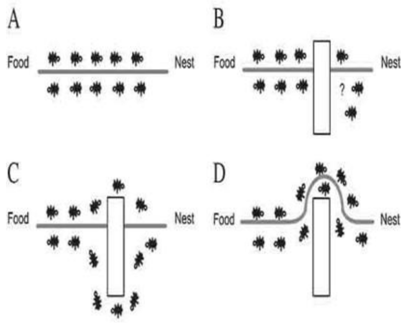
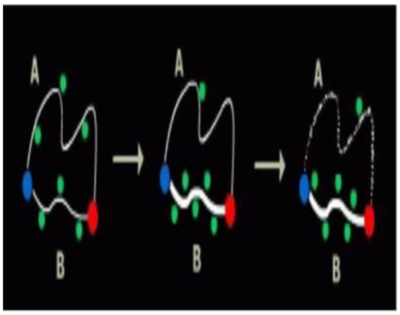


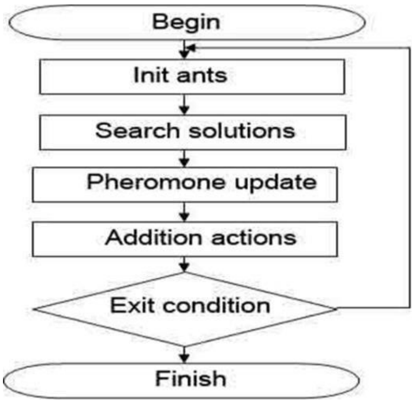
Figure-2: ACO Basic



2.2.2 Basic Ant Algorithm Operation

The basic principle of the algorithm is to have a population of artificial ants that cyclically construct solutions to a combinational optimization imposes a definition of the optimization problem into graph .In which the ants move along every branch from one node to another node and so construct paths representing solutions. Starting in an initial node, every ant chooses the next node in its path according to trail update and state transition rule.

Figure-3: Steps of ACO Algorithm



2.2.3 ACO for Feature Selection

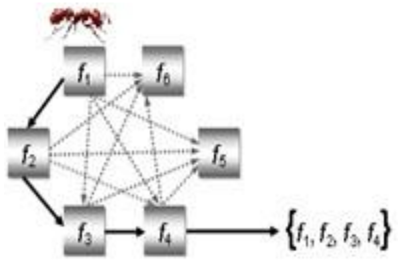
The feature selection task may be reformulated into an ACO-suitable problem. ACO requires a problem to be represented as a graph-here node represent features, with the edges between them denoting the choice of the next feature. Its a main idea of different modifications of ACO method to solve feature selection task. Four modifications of ACO method for feature selection are proposed:

- Finding feature set
- Modification based on representation of nodes by features
- Modification based on representation of nodes by informative ness

Steps to Solve a FSS (feature subset selection) problem Using ACO Algorithm. The FSS search space (features) is represented by a weighted graph (nodes with edges connecting them), where the nodes represent features and the edges denote the choice to select the next features. An optimal subset of features can be searched by an ant that walks through the graph where a predefined number of nodes (features) are visited (selected) that satisfies a traverse stopping criterion. The pheromone trail updating is defined to lay an amount of pheromone proportional to the quality of the best solutions achieved. On the other hand, pheromone trail updating is defined to lay more pheromone values for all the features in any high quality solution whose performance effectiveness is better than a predefined effectiveness value. The heuristic information is statistically defined by chi square scores for the features in the search space. Choice of the Chi-square is influenced by its performance effectiveness.

An effective Meta heuristic algorithm must achieve an appropriate balance between exploitation of the search experience gathered so far and the exploration of\ the unvisited search space. A simple approach in the balance of exploration and exploitation is tuning α and β, where determines the influence of the pheromone trail and determines the effect of heuristic information (Chisquare).Because of the huge number of features in TC tasks, feature candidate lists are created using chi-square statistic scores, we have selected a number of top features (the number of features in each candidate list is less than or equal 10 times the number of features in each feature subset). Feature selection is one of the applications of subset problems (SSP). Given a feature set of size n, the FS problem is to find a minimal feature subset of size s while retaining a suitably high accuracy in representing the original features. Therefore, there is no concept of path. A partial solution does not define any ordering among the components of the solution, and the next component to be selected is not necessarily influenced by the last component added to

Figure-4: ACO for Feature Selection



3. Problem Statement

As we know that "*A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other".* So after studying and analyzing both GA and ACO algorithms the authors found that in the existing algorithms it is difficult to maximize the over all information gain associated with a feature subset and the mutual co-relation between the among the feature is not fully minimized.

4. Proposed Model

This section proposes the multi-objective criteria to address the problem of feature subset selection. The competing objectives considered in this work are:

- Maximization of the overall information gain associated with a feature subset.
- Minimization of the mutual correlation among the feature.
- Reduction in size of the feature subset.
-

The goal of the proposed work is to find a number of small subsets of highly informative but nonredundant features. Here binary encoding has been used for representing the feature subset. Each chromosome is represented as a sequence of 0s and 1s. The bit value 1represents the presence of a feature whereas a 0 bit represents its absence.

Here, we have taken the trimmed mean of information gains to avoid the influence of extreme values or outliers as per the equation given below.

$$f_1(X) = \sum_{i=1}^{|X|} x_i \; where \; x_{i=} \begin{pmatrix} 1 + IG(x_i), \; IG(x_i) < m \\ -1 - IG(x_i), \; otherwise \end{pmatrix} \text{.......................... eq-1}$$

X represents the whole feature subset, xi signifies the ith feature of the feature subset, |X| stands for the cardinality of the feature subset and m corresponds to the Trimmed mean of information gains of all the features in the data set. Here, we have taken the trimmed mean of information gains to avoid the influence of extreme values or outliers.

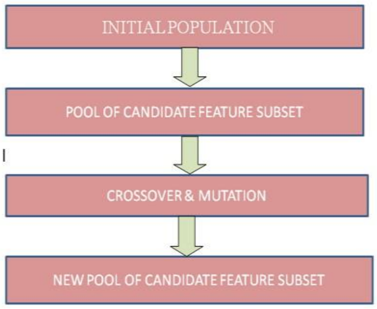
$$f_2(X) = \frac{1}{MC(X)} \text{................................eq-2}$$

$$MC(X) = \frac{1}{c(|X|,2)} \sum_{i=1}^{|X|} \sum_{j=1}^{|X|} |corr(x_i, x_j)| \text{................................eq-3}$$

$$f_3(X) = n - |X| \text{----------------------------------eq-4}$$

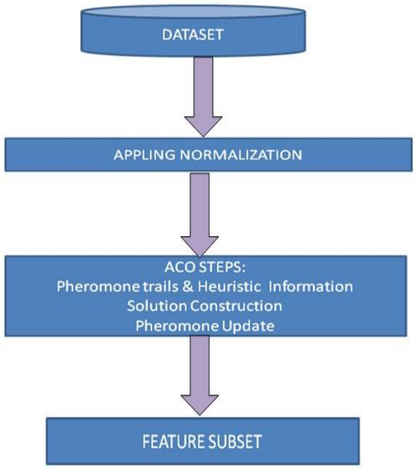
Second objective function: Maximizing non-redundancy: This objective is to minimize intracorrelation among the predicting attributes (i.e. mutual correlation).

Figure-5: GA for feature selection



In this figure we have proposed our generalized algorithm for GA based feature selection method and the results are analyzed further by using certain data sets in the experimental part.

Figure-6: ACO for feature selection



In this figure we have proposed our generalized algorithm for ACO based feature selection method and the results and graphs are analyzed further by using certain data sets in the experimental part.

5. Experimental Analysis

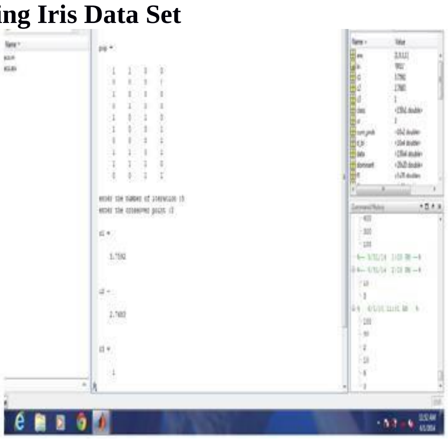
The experimental work has been carried out on Intel core processor running at 3.30 GHz with 3 GB of RAM and Windows 7 Professional as the operating system. We have used the MATLAB environment for implementing ACO and GA with the proposed multi-criteria objective function. The proposed approach has been validated on Iris dataset taken from the UCI Machine Learning Repository IRIS DATASET. The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by Sir Ronald Fisher (1936) as an example of discriminate analysis. The data set description are shown in the table below.

Table-1: **Data set description**

OPTION	IRIS	LUNG CANCER
DATA TYPE	MULTIVARIATE	MULTIVARIATE
ATTRIBUTE TYPE	REAL	INTEGER
NUMBER OF INSTACES	150	32
NUMBER OF ATTRIBUTES	4	56

Figure-7: ACO for feature selection using Iris Data Set

Figure-8: GA for feature selection using Iris Data Set



In the above figure 7 and 8 it has been analyzed that for iris data set having 150 numbers of instances the total amount of the feature selection becomes equal at a certain point and time required to compute the features is better in case of GA as compared to the ACO. If we consider big number of instances and bigger data sets ACO seems to be better in terms of feature selection as compared to GA.

Figure-9: GA for feature selection using Lung Cancer Data Set

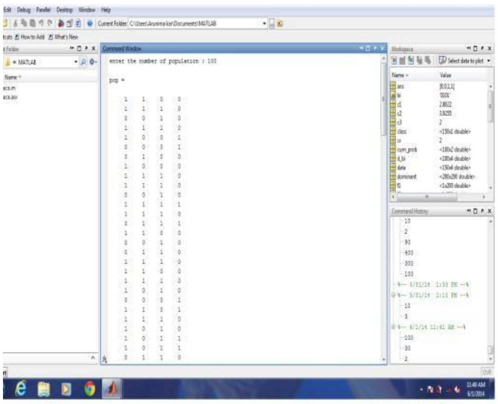
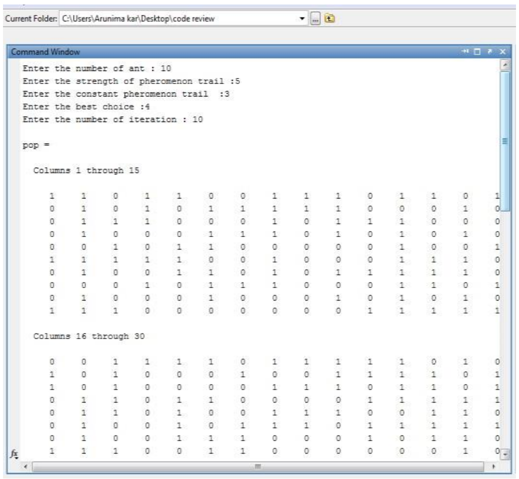
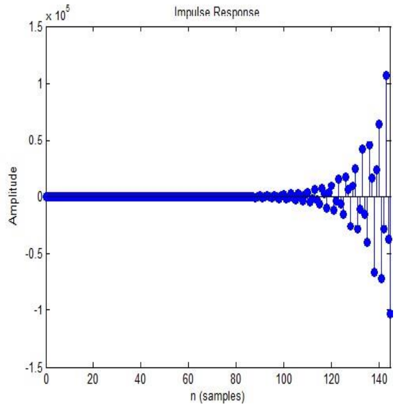


Figure-10: ACO for feature selection using Lung Cancer Data Set

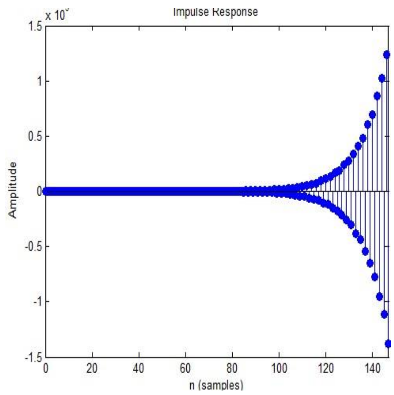

```
Current Folder: C:\Users\Arunima kar\Desktop\code review
Command Window
Enter the number of ant : 10
Enter the strength of pheromenon trail :5
Enter the constant pheromenon trail  :3
Enter the best choice :4
Enter the number of iteration : 10

pop =

  Columns 1 through 15

     1     1     0     1     1     0     0     1     1     1     0     1     1     0     1
     0     1     0     1     0     1     1     1     1     1     0     0     0     1     0
     0     1     1     1     0     0     0     1     0     1     1     1     0     0     0
     0     1     0     0     0     1     1     1     0     1     0     1     0     1     0
     0     0     1     0     1     1     0     0     0     0     0     1     0     0     1
     1     1     1     1     1     0     0     1     0     0     0     1     1     1     0
     0     1     0     0     1     1     0     1     0     1     1     1     1     1     0
     0     0     0     1     0     1     1     1     0     0     0     1     1     0     1
     0     1     0     0     0     1     0     0     0     1     0     1     0     1     0
     1     1     1     0     0     0     0     0     0     0     1     1     1     1     1

  Columns 16 through 30

     0     0     1     1     1     1     0     1     1     1     1     1     0     1     0
     1     0     1     0     0     0     1     0     0     1     1     1     1     0     1
     1     0     1     0     0     0     0     1     1     1     0     1     1     0     1
     0     1     1     0     1     1     0     0     0     0     1     1     1     1     1
     0     1     1     0     1     0     0     1     1     1     0     0     1     1     0
     0     1     0     0     1     0     1     1     1     0     1     1     1     1     1
     0     1     0     0     1     1     1     0     0     0     1     0     1     1     0
     1     1     1     0     0     1     1     0     0     0     0     0     0     1     0
```

In the above figure 9 and 10 we have analyzed the feature selections for both GA and ACO by using lung cancer data sets. It has been analyzed that ACO performs comparatively better as compared to GA. The total number of features selected by ACO is found to be more as compared to GA and that too within fewer amounts of time and less complexity. Below shown are the graphs generated by the both ACO and GA with respect to the time.

Grah-1: Feature Selection Graph for ACO



Grah-2: Feature Selection Graph for GA



The above are represented graphs for both ACO and GA for feature selection by using iris dataset and lung cancer data sets. Here X-axis represented the samples and y axis represented the amplitude. The impulse responses are hence analyzed graphically. The graph of ACO seems to be better as compared to GA in terms of feature selection.

6. Conclusion and Future Work.

Feature subset selection as a common technique used in data preprocessing for pattern recognition, machine learning and data mining, has attracted much attention in recent years. Due to the development of information acquirement and storage, tens, hundreds, or even thousands of features are acquired and stored in databases for some real-world applications. With a limited amount of training data, an excessive amount of features may cause a significant slowdown in the learning process, and may increase the risk of the learned classifier to over fit the training data because irrelevant or redundant features confuse learning algorithms. It is desirable to reduce data to get a smaller set of informative features for decreasing the cost in measuring, storing and transmitting data, shortening the process time and leading to more compact classification models with better generalization.

The main idea of ACO is to model a problem as the search for a minimum cost path in a graph. Algorithms that tend to solve problems using ACO create a search space with nodes and design a procedure to find a solution path. It generates artificial ants and works in an iterative fashion to determine solution for the problems through their indirect communication via synthetic pheromone. One approach in the feature selection area is employing population-based optimization algorithms such as Genetic Algorithm (GA)-based method and Ant Colony Optimization (ACO) based method. ACO algorithm is inspired of ant's social behavior in their search for the shortest paths to food sources. Most common techniques for ACO-Based feature selection use the priori information of features. The future work can be extended by taking more number of datasets and getting appropriate results.

7. References

[1]. H. Liu, L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering", IEEE Trans. Know l. Data Eng vol.17 (4) pp.491-502, 2005.

[2]. M.Dorigo and T.,"Stutzle,Ant Colony Optimization, The MIT Press, pp.21-22, 2004.

[3]. N. Jain and J. P. Singh, "Modification of Ant Algorithm for Feature Selection", International conference on control, automation, communication and energy conservation, 2009.

[4]. S. Kashef, H. Nezamabadi-pour, "A new feature selection algorithm based on binary ant colony optimization", Conference on Information and Knowledge Technology (IKT),2013

[5]. F. Tan, X. Fu, Y. Zhang, and A. G. Bourgeois, A genetic algorithm-based method for feature subset selection, Soft Computing, vol. 12, no. 2, pp. 111-120, (2008).

[6]. M. E. ElAlami, A _lter model for feature subset selection based on genetic algorithm, Knowledge-Based Systems, vol. 22, no. 5, pp. 356-362, (2009).

[7]. Yang and V. Honavar, Feature subset selection using a genetic algorithm, Intelligent System and Their Applications IEEE, vol. 13, no. 2, pp. 44-49, (1998).

[8]. 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. 491502, (2005).

[9]. T. M. Hamdani, J.-M. Won, A. M. Alimi, and F. Karray, Multi-objective Feature Selection with NSGA II, Adaptive and Natural Computing Algorithms, Springer Berlin Heidelberg, pp. 240-247 (2007).

[10]. A.Freitas and C. A. Kaestner, Attribute selection with a multi-objective genetic algorithm, Proc. of the 16th Brazilian Symposium on Artificial Intelligence, (2002).

[11]. C.-M. Wang and Y.-F. Huang, Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data, Expert Systems with Applications, vol. 36, no. 3, pp. 5900-5908, (2009).

[12]. G. L. Pappa, A. A. Freitas, and C. A. Kaestner, A multi-objective genetic algorithm for attribute selection, Proceedings of the 4th International Conference on Recent Advances in Soft Computing (RASC-2002), pp. 116-121, (2002).

[13]. C. Emmanouilidis, A. Hunter, and J. MacIntyre, A multi-objective evolutionary setting for feature selection and a commonality-based crossover operator, Proceedings of the 2000 Congress on Evolutionary Computation, vol. 1, pp. 309-316, (2000).

[14]. L. S. Oliveira, M. Morita, and R. Sabourin, Feature Selection for Ensembles Using the MultiObjective Optimization Approach, Multi-Objective Machine Learning, Y. J. Dr, Ed. Springer Berlin Heidelberg, pp. 49-74, (2006).

[15]. M.Venkatadri and K.Srinivasa Rao, A Multi-objective Genetic Algorithm for Feature Selection in Data Mining, International Journal of Computer Science and Information Technologies, vol. 1, no. 5, pp. 443-448, (2010).

[16]. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multi-objective genetic algorithm: NSGAII, IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182197, (2002).

[17]. A. Konak, D. W. Coit, and A. E. Smith, Multiobjective optimization using genetic algorithms: A tutorial, Reliability Engineering and System Safety, vol. 91, no. 9, pp. 992-1007, (2006).

[18]. M. Dorigo and C. Blum, Ant colony optimization theory: A survey, Theoretical Computer Science, 344, pp. 243-278, 2005.

[19]. M. Dorigo, Optimization, Learning and Natural Algorithms, Ph.D. dissertation, Department of di Electronica, Politecnico di Milano. Italy, 1992.

[20]. M. Dorigo, Learning by Probabilistic Boolean Networks, in Proc. IEEE International conf. Neural Networks, pp. 887-891, 1994.