RESEARCHERID
THOMSON REUTERS

[Kaur, 5(8): August 2018]
DOI- 10.5281/zenodo.1341054

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

EVOLUTIONARY METAHEURISTIC ALGORITHMS FOR FEATURE SELECTION: A SURVEY

Sandeep Kaur[*1] & Vinay Chopra[2]

[*1]Research Scholar, Computer Science and Engineering, DAV Institute of Engineering and Technology, India

[2]Assistant Professor, Computer Science and Engineering, DAV Institute of Engineering and Technology, India

ABSTRACT

The advancement of high-throughput technologies has resulted in exponential growth of high dimensional data. It has been a challenge to efficiently find patterns and discover knowledge using data mining and machine learning techniques from the collected data. However, this data contains redundancy and irrelevant features. Therefore, dimensionality reduction techniques such as feature extraction and feature selection are applied to deal with the data that is associated with noise. There are several feature selection techniques that minimize redundancy. However, more promising results can be attained using metaheuristics algorithms that are based on natural evolution. Through this paper, we overview and highlight the importance of metaheuristics and surveys existing feature selection algorithm for data mining.

Keywords: *Feature Selection, Particle Swarm Optimization, Genetic Algorithm, Ant Colony Optimization*.

I. INTRODUCTION

Feature Selection is a major research topic for the development of classification methods. FS (subset selection or attribute selection) is the automatic selection of attributes in dataset that are most relevant for application of a learning algorithm[18] . The best subset is selected that contains least number of dimensions contributing well to the accuracy of the learning classifier. Hence, rest of the features and irrelevant dimensions are discarded and only best ones are kept. FS is usually done in the pre-processing stage and is very efficient technique to deal with noisy and redundant features and tend to keep only those features which are best suited for the classifier. There are basically two methods by which FS can be carried out:

1. Forward Selection: In this type of selection, initially there are no variables and variables are added gradually at each step so as the error is decreased. This process halts when further addition does not decrease the error rate.
2. Backward Selection: This is contrasting to the prior one. In this case, the selection starts with considering all the variables and removing one by one at each step until any further removal increases the error considerably.

A typical feature selection process consists of four basic steps as shown in Figure 1. The first step is subset generation which is a search procedure that produces candidate feature subsets for evaluation. After that, each candidate subset is evaluated and compared with the previous best subset according to a certain evaluation criterion. If the new subset is found to be better than the previous one, previous best subset is replaced. The process of subset generation and evaluation continues repeating until a given stopping criterion is satisfied. Finally, result is validated by prior knowledge[16] . Feature selection has successfully been applied in many fields such as classification, clustering, association rules and regressions. FS algorithms broadly fall into three main categories: the filter approach, the wrapper approach and the hybrid approach. In the filter approach [17] , feature subsets are selected and evaluated without requiring any classifier. In case of wrapper approach [8] , one predetermined classifier's performance is used as evaluation criterion. In this approach, features are selected in such a way that subset improves the performance of the classifier. However, wrapper approach is proved to be more computationally expensive than the filter approach. On the other hand, hybrid approach considers taking advantage of the two approaches by exploiting their different evaluation criteria in different search stages

This paper attempts to review the field of feature selection based on most intensively used metaheuristics algorithms in the field of data mining such as Particle Swarm optimization(PSO), Genetic Algorithm (GA) and Ant Colony Optimization (ACO). All these listed algorithms are very promising and researchers will gain an insight of incorporating these algorithms in their task. This survey aims at highlighting the use of these algorithms for feature selection and also work of some researchers has been presented.

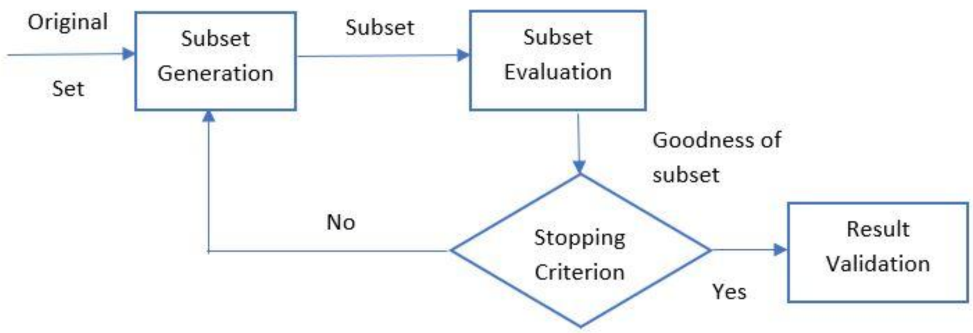

Figure 1 Key Steps in Feature Selection

II. FEATURE SELECTION TECHNIQUES

1. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is recently gathering much research attention in the field of machine learning. PSO is based on the social behavior associated with bird flocking for optimization problems, firstly developed by Kennedy and Eberhart [9]. This swarm intelligence based technique is computationally less expensive and converge more quickly. Therefore, it has been used as an efficient technique is many fields, including feature selection. In PSO, each solution is represented as a particle in the swarm where each particle is represented by a vector. Algorithm initializes with n-randomly chosen particles and finds the optimal particle iteratively. Therefore, each particle has some velocity. Also, according to experience and that of neighbors, each particle updates its position and velocity. The best previous position of the particle is called as personal best and the best position obtained by the population of particles is called as global best. Thus, velocities and positions of each particle are updated to find the optimal solution based on personal best and global best. At last, the algorithm stops when a predefined criterion is met such as best fitness value or maximum number of iterations (Figure 2).

Many researchers have recently focused much on PSO for solving feature selection problems. PSO can overcome the limitations of other feature selection approaches such as stochastic hill climbing. Wang et al. [10] propose a feature selection algorithm based on improved PSO and rough set theory since rough set can handle uncertainty, imprecision and ambiguity. This paper also presents that a drawback of rough set when used with stochastic hill climbing; it can, however, achieve better and promising results when used with improved binary PSO. Liu et al. [11] design a modified multi-Swarm PSO which consists of a number of sub-swarms and a scheduling module. Moreover, in order to describe the competition among the swarms, the mechanism for survival of the fittest is introduced in which four rules are designed to decide whether a sub-swarm should be destroyed or reserved. In this, the scheduling module monitors and controls each sub-swarm according to the rule during the iterations of the algorithm. Chung-Jui Tu et al. [12] implement PSO with SVM with one-versus-rest method serve as a fitness function of PSO. This method is applied to five classification problems and gives better and higher classification accuracy compared to other feature selection approaches. Moreover, this paper shows that suitable parameter adjustment enables PSO to increase the efficiency of FS. Li-Yeh Chuang et al. [1] exploit an improved binary PSO for feature selection and the K-nearest neighbor (K-NN) for gene expression data. In the task of classification of gene expression data samples, feature selection is most crucial in order to analyze gene profiles correctly due to the fact that only a small number of genes show a strong correlation with a certain phenotype as compared to the total number of genes exploration. In PSO, each particle adjusts its position according to two fitness value, pbest (local fitness value) and gbest (global fitness value), to avoid being trapped in a local optimum. Thus, superior classification results will be prevented if the gbest value gets trapped in a local optimum. Therefore, the improved PSO approach [] aims at retiring gbest under such conditions and thus resetting gbest to achieve superior results. Azvedo et al. [2] presents a feature selection model using PSO as the optimization algorithm and SVM as the verification algorithm for the task of keystroke dynamic systems. A PSO variation is created where each particle is represented by a vector of probabilities that indicate the possibility of selecting a particular feature and it directly affects the values of other features. This approach is able to reduce classification error, processing time and feature reduction rate as compared to GA.

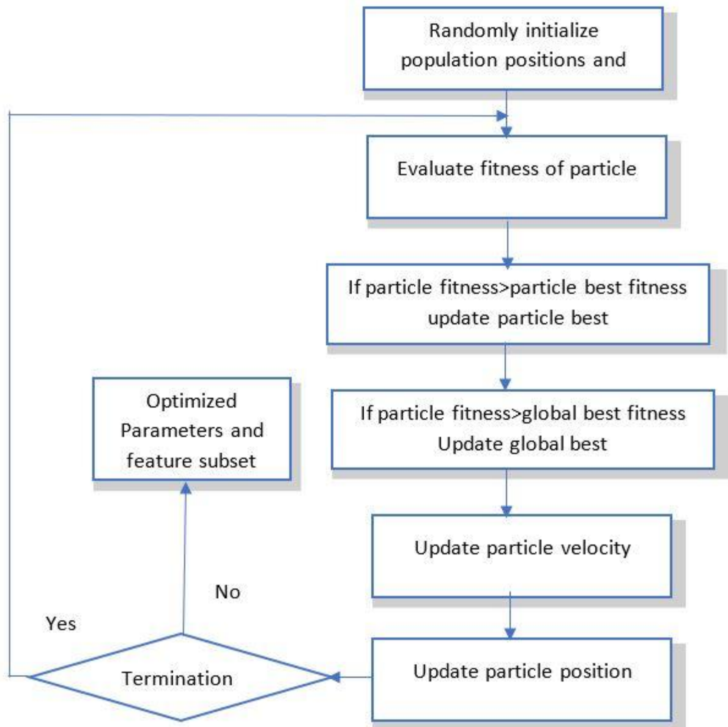

Figure 2. Flowchart of Particle Swarm Optimization

2. Genetic Search

Genetic Algorithms (GAs) are optimization algorithms that work on Darwinian models of population biology. GAs are capable of finding near-optimal solution for multivariable functions. Moreover, GAs are adaptive search techniques which have represented significant improvement over number of random and local search methods [15].

GA being an iterative algorithm operates on initial population which is set of candidate solutions. Each solution is represented as a chromosome and vice versa [13]. Population is randomly generated in the initial step and every individual in the population is assigned a fitness value by means of a fitness function that determines its quality. Moreover, a chromosome is evaluated by a fitness function that reflects the quality of the solution. The input of the fitness function is the chromosome and the output is the fitness value of this chromosome. In each iteration, fitness of each candidate solution is calculated. In the next iteration, another population is generated where the fittest individuals are more likely to survive than less fit individuals. These selected individuals act as parents for the next generation which is achieved by applying reproductive operators like crossover and mutations. The flowchart of Genetic Algorithm is shown in Figure 3.

The main thing to consider in applying GAs to any problem is selecting an appropriate representation and an adequate evaluation function. GAs are very effective in resolving the high dimensionality problem as in [3]. Uguz [3] proposed two-stage feature selection and feature extraction to improve the performance of text categorization. Moreover, it has been asserted in this paper that it is highly important to rank the features to effectively deal with high dimensionality of data in document categorization. Therefore, three different algorithms have been applied in the proposed approach. In the first stage, each term is ranked depending on its importance for classification in decreasing order using Information Gain (IG). Next, Principle Component Analysis (PCA) selected for feature selection and GA for Feature Extraction are applied separately to the terms of highest importance. Although terms if high importance are acquired by IG method, the main problem of high dimensionality still is the main problem which is resolved by adopting GA in this methodology.

GAs are proved to optimize the output of Artificial Neural Network (ANN) based systems as well. ElAlami [4] described a novel algorithm for feature subset selection based on Genetic Algorithm to optimize the output nodes of trained ANN. In this approach, ANN is first trained on the input features and the corresponding class where each input unit corresponds to a single feature and each output unit corresponds to a class value. GA is used to find the optimal values of input features for each output node, which maximizes the output function by reducing the dimensionality. The performance of methodology is tested on two different datasets and reduces the dimensionality by 50% and 33.3% which yields a good indication of algorithm stability.

GAs have been very effective in improving the accuracy of the classifiers in filed of medical diagnosis of a disease by selecting only relevant features that contribute to the optimal solution. M. Anbarasi et al. [5] predicted the presence of heart disease with reduced attributes more accurately by incorporating GA as feature selection algorithm. The use of Genetic algorithm selected those features that contribute towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken. GA reduced thirteen attributes to six attributes that were fed into three different classifiers such as Naïve Bayes, Clustering and Decision Tree.

ISSN 2348 – 8034
Impact Factor- 5.070

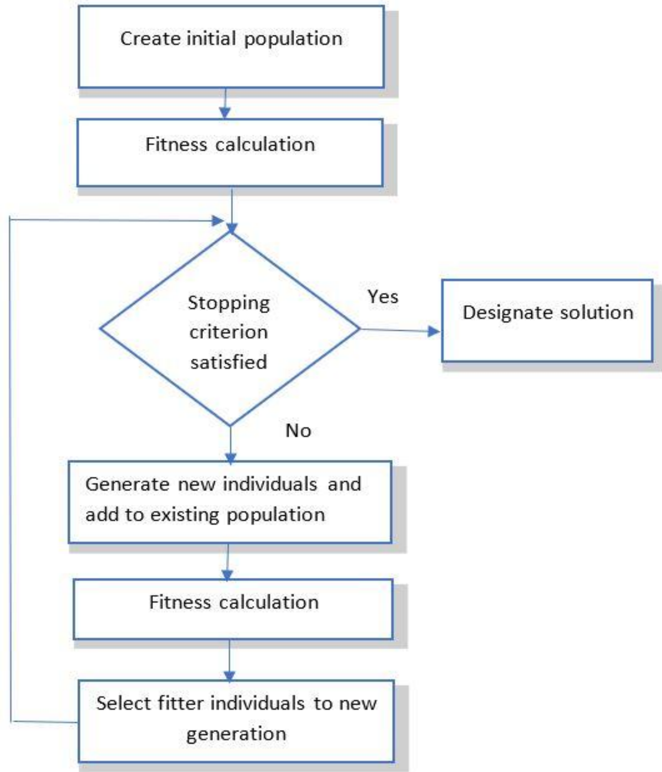

Figure 3. Flowchart of Genetic Algorithm

3. Ant Colony Optimization

Since feature selection is an NP-Hard problem, heuristic algorithms which are based on natural behavior of living beings are very promising and provide better accuracy in many classification tasks such as pattern classification, sentiment analysis, text mining and many areas related to medical. In recent years, Ant algorithms also have been applied to the problem of feature selection. Ant colony optimization was first put forth by Marco Dorigo [14] which was initially applied to solve travelling salesman problem.

To reformulate ant colony optimization into a feature selection problem, it is required to represent the problem as a graph. In this task, features are represented as nodes and edges between the nodes that denote the choice of the next feature. Nodes are fully connected to let other features to be selected. The optimal feature subset is selected according to the traversal of an ant where the minimum nodes are visited. In constructing the solutions, each ant is assigned an initial node(feature) randomly. Tour is constructed by applying probabilistic transition rule which is based on heuristics information and pheromone trails. When the stopping criterion is satisfied, the ant terminates its traversal and outputs the best feature subset (Figure 4).

An intensive research has been done on feature selection using Ant Colony Optimization in the field of text mining . ACO improves classification accuracy by selecting only relevant features as in [6]. M.F. Zaiyadi [6] proposed a novel hybrid approach for feature selection in text document categorization based on ACO and Information Gain (IG). The main motive of combining these two algorithms is to overcome the drawback of IG. IG is less effective when operated with highly redundant features. However, ACO is proved to provide good solutions when dealing with high dimensionality data. Ahmed Al-Ani [7] exploit a novel feature subset selection using ACO metaheuristic. A hybrid evaluation measure combining wrapper and filter approach of feature selection is used in this work that is able to estimate the overall performance of the subsets as well as the local importance of the features. The

ISSN 2348 – 8034
Impact Factor- 5.070

performance of subsets is estimated by classifier performance which is wrapper evaluation function. However, Mutual Information (MI) measures the local importance of a given feature being a filter evaluation function. The strength of using hybrid feature evaluation approach over GA-based FS is reduced computational cost because using wrapper approach alone requires far more computational cost. Abd-Alsabour et al. [8] applied wrapper based feature selection approach based on ACO in the field of pattern classification. This approach deals with the problem of feature selection as a binary problem where a set of binary bits is associated with each ant to represent its feature subset selection where if n^{th} bit is 1 feature number n in the dataset is selected, otherwise feature is not selected. Therefore, in the initial step bits are randomly initialized to zero's and one's. At each construction step, ant select the next feature according to the heuristic information and pheromone value. Then each solution by ants are passed to a classifier such as support vector machines to calculate its accuracy so that best and optimal feature can be calculated. The algorithm halts after reaching a predefined set of iterations.

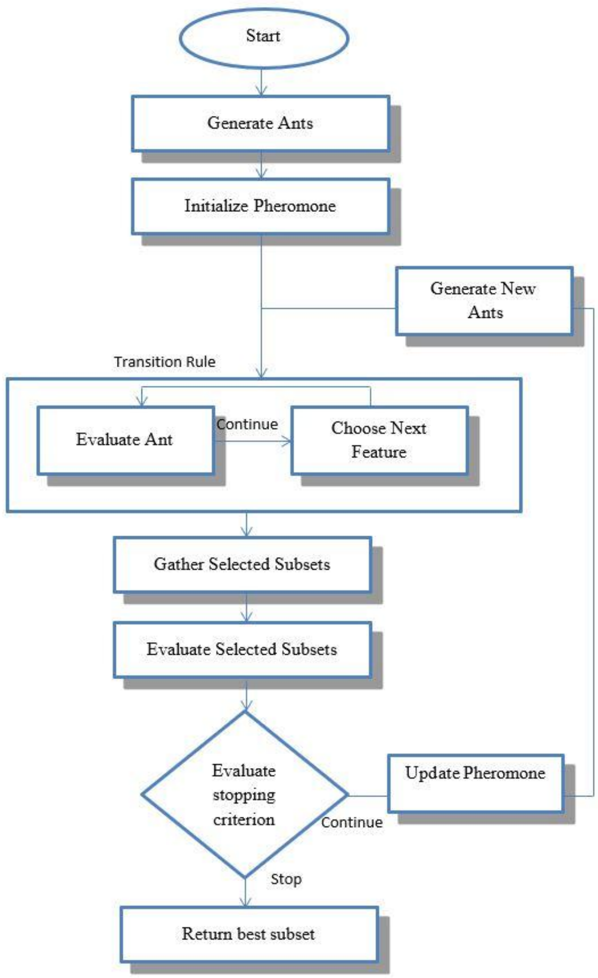

Figure 4. Flowchart of Ant Colony Optimization

III. CONCLUSION

In this paper we have provided an overview of feature selection techniques based on metaheuristics. The literature on feature selection is very vast which encompass the applications of data mining and machine learning applications. Feature selection techniques reveal that it is very crucial to remove noise from high dimensional data as it may affect the performance of the classifier. Evolutionary algorithms such as Particle Swarm , Genetic and Ant colony optimization can be selected for various tasks in data mining.

IV. ACKNOWLEDGEMENTS

This research was supported by DAVIET .We thank our faculty members who provided insight and expertise that greatly assisted the research. We thank [Dr. Manoj Kumar, Principal] for his assistance We would also like to show our gratitude to the HOD Of the department for sharing her pearls of wisdom with us during the course of this research.

REFERENCES

1. *Chuang, L. Y., Chang, H. W., Tu, C. J., & Yang, C. H. (2008). Improved binary PSO for feature selection using gene expression data. Computational Biology and Chemistry, 32(1), 29-38.*
2. *Azevedo, G. L., Cavalcanti, G. D., & Carvalho Filho, E. C. (2007, September). An approach to feature selection for keystroke dynamics systems based on PSO and feature weighting. In Evolutionary Computation, 2007. CEC 2007. IEEE Congress on (pp. 3577-3584). IEEE.*
3. *Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. Knowledge-Based Systems, 24(7), 1024-1032.*
4. *ElAlami, M. E. (2009). A filter model for feature subset selection based on genetic algorithm. Knowledge-Based Systems, 22(5), 356-362.*
5. *Anbarasi, M., Anupriya, E., & Iyengar, N. C. S. N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology, 2(10), 5370-5376.*
6. *Zaiyadi, M. F., & Baharudin, B. (2010). A proposed hybrid approach for feature selection in text document categorization. World Academy of Science, Engineering and Technology, 48, 111-116.*
7. *Al-Ani, A. (2005). Feature subset selection using ant colony optimization. International journal of computational intelligence.*
8. *Abd-Alsabour, N., & Randall, M. (2010, December). Feature selection for classification using an ant colony system. In e-Science Workshops, 2010 Sixth IEEE International Conference on (pp. 86-91). IEEE.*
9. *Eberhart, R., & Kennedy, J. (1995, October). A new optimizer using particle swarm theory. In Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on (pp. 39-43). IEEE.*
10. *Wang, X., Yang, J., Teng, X., Xia, W., & Jensen, R. (2007). Feature selection based on rough sets and particle swarm optimization. Pattern recognition letters, 28(4), 459-471.*
11. *Liu, Yuanning, Gang Wang, Huiling Chen, Hao Dong, Xiaodong Zhu, and Sujing Wang. "An improved particle swarm optimization for feature selection." Journal of Bionic Engineering 8, no. 2 (2011): 191-200.*
12. *Tu, C. J., Chuang, L. Y., & Yang, C. H. Feature Selection using PSO-SVM.*
13. *Vafaie, H., & De Jong, K. (1992, November). Genetic algorithms as a tool for feature selection in machine learning. In Tools with Artificial Intelligence, 1992. TAI'92, Proceedings., Fourth International Conference on (pp. 200-203). IEEE.*
14. *Dorigo, M., Birattari, M., Blum, C., Clerc, M., Stützle, T., & Winfield, A. (Eds.). (2008). Ant Colony Optimization and Swarm Intelligence: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008, Proceedings (Vol. 5217). Springer.*
15. *Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. (2000). Dimensionality reduction using genetic algorithms. IEEE transactions on evolutionary computation, 4(2), 164-171.*
16. *Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(3), 131-156.*
17. *M. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering – a filter solution. In Proceedings of the Second International Conference on Data Mining, pages 115–122, 2002.*
18. *A.L. Blum and P. Langley. Selection of relevant features and examples in machine learning. Artificial Intelligence, 97:245–271, 1997.*