J. Basic. Appl. Sci. Res., 8(2)1-14, 2018

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Critical Analysis of Feature Subset Selection Using Soft Computing Based Techniques

Arslan Ellahi[1*] and Waseem Shahzad[2]

[1]MY University, Islamabad, Pakistan
[2]National University of Computer and Emerging Sciences, Islamabad, Pakistan

Received: November 2, 2017
Accepted: January 23, 2018

ABSTRACT

Feature subset selection is a vital issue in solving the complex problems of data mining, related to the performance of data. Data sets are very large, contains relevant as well as irrelevant information. Irrelevant features increase complexity and time of processing, so extraneous feature should be wiped out by some algorithmic measure to improve the computation time, predictive accuracy and performance. Soft computing techniques are useful for finding the optimal solution of different problems. In this paper, we critically analyze the different soft computing techniques applied to feature subset selection i.e. genetic algorithm, ant colony optimization, particle swarm optimization, differential evolution, and rough sets. We thoroughly investigate the strengths and weaknesses of these techniques and also the research trends related to these techniques.
KEYWORDS: Feature subset selection, Soft Computing, Evolutionary Computation, Swarm Intelligence, Supervised Learning.

1.INTRODUCTION

Data is growing gigantically in the world which results in various issues related to its relevance and importance. We have to solve the issue related to irrelevant information and redundant data to be eliminated from dataset [1]. Feature selection is a vital technique used for improving the performance of a learning system. Knowledge discovery from data is a difficult and expensive task, and we cannot do the processing of complete data as it is entirely impossible in most of the cases [5].
Reducing no of features is not a trivial task, for this, we are required data mining techniques to improve the predictive accuracy of a learning model and also reduce the computational complexity. Redundant and irrelevant features only increase the complexity of data and also increase the processing time and cost [3], [6], [7].
Large dimensional datasets usually take alonger time to search due to which they have high computational complexity. Most of the real world problems have high dimensional datasets [1]. Selecting an optimal number of features is a hot research area. The objective of feature selection is to find useful features that enhance the performance and predictive accuracy. Relevancy is very important in feature selection, relevant and significant feature are useful in knowledge extraction, and they are used for optimization and enhancement the model. There are three major categories of feature subset selection i.e. *filter based model*, *wrapper based model* and *hybrid model* [2]. Soft computing techniques gain popularity to find out the optimal solution to a problem. There have been various techniques of soft computing.
Data mining process has some steps to reach towards knowledge discovery. It is a complex task to transform huge amount of data into refined knowledge and to use it for decision making. One of the challenging problems is to transform this huge amount of data into an accessible and actionable knowledge. Patterns are the most useful knowledge that has to be extracted from the data. Domain experts utilize this knowledge for decision making.
In this paper, we analyze the different feature subset selection techniques based on soft computing [3], [6]. Soft computing techniques are used to solve the complex optimization problems, and data mining and soft computing can be used collectively to extract knowledge from the datasets. Many problems of data mining have been solved using these techniques. In literature, it has been showed that these techniques find out the useful features from the data set to build a good learning model [6]. In this paper, we thoroughly investigate and analyze the various soft computing techniques proposed for feature subset selection.

2. Feature Selection

Feature subset selection is used in many meadows e.g. bioinformatics, pattern recognition, machine learning, and majorly in the field of data mining [3], [6], [7], [15], [16], [18]. Dimensions of datasets are very important and irrelevant, or redundant dimensions are useless; to improve predictive accuracy and performance useless feature must be eliminated. Feature selection focuses on selecting highly accurate features that are neither irrelevant nor redundant. This result improves computational time and data storage [3]. [6]. Curse of dimensionality is the core focus of feature subset selection.

Corresponding Author: Arslan Ellahi, MY University. Islamabad, Pakistan.
Email: arslan.ellahi@myu.edu.pk

A feature that provides useful information or it help in improving the predictive accuracy of the model should be selected. Overfitting occurs when irrelevant features build the model; they also lead towards miss classification [2], [3].

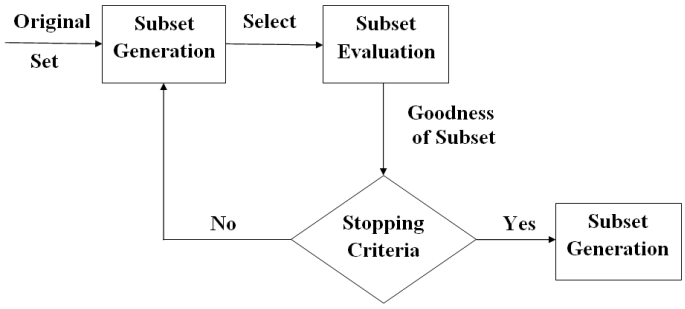

Fig.1. Three main ingredients of feature selection method

A feature that provides useful information or help in improving the predictive accuracy of the model should be selected. In this context, the objectives are:

- *To use optimal searching for feature subset selection that takes less amount of time and generates an optimal number of features.*
- *To increase the predictive accuracy and improve classifiers performance by using an optimal number of features.*

There are many approaches used to select features from the dataset. There are three main types of feature subset selection i.e. filter based, wrapper based and hybrid feature subset selection. In next section we will these in detail [18].

2.1 Filter Based Model

High dimensional dataset needs optimization and reduction of redundant features which usually increase computational search time and predictive performance [5], [6]. There are many approaches on the basis of which we can select our feature from dataset. Feature selection is composed of two main techniques that are filter based method, wrapper based method and on technique is combination of both that is called as hybrid technique for feature subset selection, this technique contain the taste of both filter and wrapper based approach. Below we have discussed all the techniques [3].

In filter based model first of all subset of features are extracted from major features before applying actual learning algorithm of feature selection [8]. Best subset features are extracted using some evaluations and iteration in one pass; they are extracted on the basis of some independent evaluation criteria. This can give speed, filter based technique is computationally less expensive and more generalized, filter based technique is independent of learning algorithms that make it more general than specific [1]. But that support cannot guarantee appropriate accuracy, greater the number of features greater will be their generality. Filter based technique depend on statistical measurements that are used to develop relationship between different features [19]. Many techniques have been applied on filter based model in order to generate efficient and vibrant results. In some of the cases it might fail to select right subset of features. Major drawback of filter based model is that at a time it only focuses on one subset of feature at a time and also it compromise on predicative accuracy.

Fig. 2. Filter based feature selection method.

2.2 Wrapper Based Model

The second technique of feature subset selection is a wrapper based model, which uses a learning algorithm during the feature extraction process and try to find out those features that improve the predictive accuracy [18]. The accuracy of wrapper based algorithm is higher than filter based method, and it is computationally expensive as compared to filter based methods. The wrapper-based method may overfit on small datasets. They used searching methods to find appropriate feature subset space [1], [2]. Figure 3 explains the process of wrapper based methods.

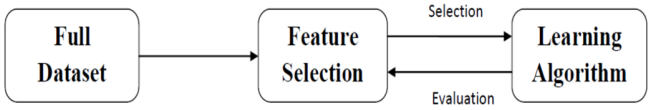

Fig. 3. Wrapper based feature selection method.

2.3 Hybrid Model for Feature Selection

The hybrid method is composed of the combination of filter based and wrapper based methods. Hybrid methods use the evaluation criteria of filter and wrapper model and it has been used for the different problems. This method is more effective as it made up from the combination of two it reduce the computational complexity of again and again training of the algorithm. The idea of the hybrid approach is to combine and absorb useful information from different feature subsets that are small in size; itavoids individual filter or wrapper algorithm for being used that will increase classification performance [18].

3. Feature Subset Selection Process

Following are the main steps involved in feature subset selection process [2], [5], [17], [18], [19], [25], [29], [35], [36].

3.1 Initialization Point

The process of feature subset selection starts with the initialization phase;there are two different strategies of initialization process one is forward feature subset selection, in which we start with empty features and keep on adding feature one by one that are selected. We continue to add features until accuracy have been improved. Second is backward feature subset selection, in which we start with a complete subset of features and then eliminate the features one by one until accuracy has been improved.

3.2 Approach for Searching

Searching is one of the most computationally complex task, in which we have to check full dataset. Exhaustive searching is one of the possible solutions but computationally very expensive, we have to explore all the features in our dimensionality space. In large dimensional datasets, this approach cannot exist for a long time and may crash due to inefficiency. We have to follow quite realistic approaches to reduce the computational complexity of our model. The quality of selected features should be high; usefully extracted features improve the predictive performance and classification accuracy. In this paper, we will explore and analyze various search based algorithms used in feature subset selection domain.

3.3 Estimation of Selected Subsets

Next phase of feature subset selection is to observe the selected features by useful information they hold. Two feature subset selection techniques will be evaluated; that are wrapper based and filter based approach by their selected subsets. Wrapper-based methods are used to select the goodness of a feature by applying different learning algorithms and evaluate by efficient performance and effective results. Wrapper-based approaches are more effective than filter based because filter based approaches do not use any learning algorithm during training.

3.4 Stopping Measures

The stopping criteria of feature subset selection algorithm can be, when performance does not improve further by increasing features or decreasing features than we should stop the algorithm. When a certain amount of threshold met then we can also stop feature subset selection algorithm.

4. Evolutionary Algorithms

4.1 Evolutionary Algorithms Overview

Nature is composed of many biological experiences, each of which has different running criteria. The idea of biological algorithms came up from the Darwin evolutionary systems of animals and plants. They are inspiredby nature and environmental behavior mostly applied on global optimization problems and also in function optimization of large dimensional data. Convergence varies from problem to problem but they are highly effective, and their predictive accuracy is high. Strategy building and strategy optimization are the core focus on which the complete evolutionary process runs. Most of the evolutionary algorithms contain some population at time t, variation operators, and selection operators [3], [4], [5], [6], [38], [41], [42].

4.2 Feature Selection Perspective

Evolutionary algorithms have been used for optimization of the feature subset selection process for improving performance and predictive accuracy. Evolutionary techniques have been applied to large dimensional datasets to select useful feature and to extract useful dimensions. The literature showed that these algorithms have been very useful for feature selection [26].

Evolutionary systems have different steps to reach the final solution. There are different evolutionary algorithms like

evolutionary strategies, evolutionary programming, genetic algorithms, genetic programming, differential evolution, co-evolution, and culturalevolution. Every evolutionary algorithm has a different effect in term of its performance, computational complexity and predictive accuracy [21].

4.3 Genetic Algorithms

Genetic algorithm (GA) is considered as randomized optimization algorithm, which uses searching techniques in its population space that is broadened across its sides. It is inspired by biological intelligence used to find difficult optimization problems. It consists of solutions as a population that contains individuals as chromosomes [41]. The basic concept of GA is the survival of the fittest.

1. Genetic Algorithm for Feature Selection

A genetic algorithm has a major contribution in the field of feature selection; itrepresents the fittest features by calculating their fitness value by some measures and evaluation criterion. If the feature has more fitnesshence it provides useful information. Survival of the fittest means that best-fit solutions or features will survive, and they will produce offspring's and the less fit chromosomes or feature that are providing useless information in the dataset will be culled from the data. Each chromosome represents a point in the population space [18].

A geneticalgorithm has found better subsets which ultimately improved the classification performance. A genetic algorithm is computationally fast and searches the high dimensional datasets very quickly and effectively. The features that are unrelated and not in use can be used and grouped together with some other features to form a subset of features. But there are some of the cases in which grouping can put some extra cost to the genetic algorithm and increase the computational complexity of the algorithm. To avoid this situation there needs some hybridization technique which can increase the combination to select the subsets of features [18].

1. Impact of the Algorithm

A genetic algorithm has two reproduction operators, crossover, and mutation that are used to create offspring from selected parents. These operators help in feature selection in an aspect that they are used to cross over or mutate two or more features with each other to produce anew subset of features. Mutation adds diversity in the population or features that are selected from the dataset. The feature subset selection process has beenimproved by using the genetic algorithm. Computational complexity is the key element in every algorithm, and genetic algorithm is very fast and generates only useful subsets [43, 46, 47, 59].

2. Current and Future Research

Tan et al. [13] proposed a technique for improvement of feature subset selection by using a genetic algorithm, in this technique they have used the microarray data that is made up of a huge number of genes or feature which contain small sample sets to make classification or prediction difficult. Many techniques of feature selection produce enormously different results on same data. The genetic algorithm improves the feature selection by combining multiple useful features so that best prediction results can be gained. The genetic algorithm helps to build the best classifier for the dataset and this technique can find feature subsets with the finest accuracy.

Huang et al. [20] proposed a technique for feature subset selection by using a hybrid genetic algorithm and focusing on the wrapper-based approach of feature subset selection. Classification is based on finding a subset of features by using two phases strategy one is optimization phase in which complete global searching has been performed for finding the optimized feature subset by using wrapper based method. Transfer of information takes place between classifiers for choosing the fitness function by using genetic algorithms. Second optimization phase is locally searched by using filter based approach and collect the required amount of information needed for feature selection. Both inner and outer optimization collaborate each other for making high performance and high efficiency. Experimentation results showed that the algorithm has achieved good performance and classification accuracy. This technique used entropy and information gain as performance measures. Some of the ambiguities remain in the idea as most of the focus is on wrapper-based approach and wrapper algorithm is computationally complex and takes a lot of time to build the final solution.

Tang et al. [21] proposed a combined technique for feature subset selection in which they have used a genetic algorithm for the feature subset selection. They have done extensive experimentation to discuss the results of the genetic algorithm, and it has been proved that this technique is effective and robust with high performance and predictive accuracy as compared to the single feature subset selection algorithm.

ElAlami et al. [25] proposed a technique for feature subset selection based on genetic algorithm and artificial neural networks. By utilizing both techniques on extracting best features, performance has been improved. GA is used to optimize the weights of the neural network.

Oh et al. [28] proposed a technique for feature subset selection by using the hybrid genetic algorithm. They have applied different techniques to improve the performance of genetic algorithm-like local search operations that are embedded in a genetic algorithm to improve the effectiveness of feature subset selection. This hybridization policy improved the searching quality and selects the optimized features. There are some parameters that have been used to improve the search quality like effective search and timing ratios.

There are some of the measures in the genetic algorithm which can be enhanced like its parameters and improvement in genetic operators so that predictive accuracy of features can be improved and best possible subsets of features can be extracted. A hybrid genetic algorithm technique can be improved by making small changes in a genetic scheme such as the arrangement of chromosomes and genetic operators. By analyzing the change in population and further training of two proposed parameters will enhance its performance and improve the predictive accuracy for selecting optimized features [27].

3. Genetic Algorithm Strengths in Feature Selection

There are many algorithms that have been appliedto feature selection, and every algorithm has different performance and predictive accuracy, genetic algorithms use the combination of different existing techniques to work on feature subset selection. GA can find the optimal subset of features to perform efficiently and to produce a good classifier.

GA has been used to extract the features from a data set and ranked the features by their relevance and predictive accuracy. There is a need to remove the redundant features and grouping of features should have particular criteria for selecting features from a pool of features. These are the two most important issues that should be improved. One possible thing in this regard is to construct the feature pool for GA and to design a fitness function according to these criterions [13].

Genetic algorithm finds the best and most accurate, optimized features for the model that result in maximizing the output function of the class. Dominating features have been selected from the input of datasets that improved the predictive accuracy. They have used car and monk's datasets to check the predictive accuracy; the proposed technique reduced the dimensionality of data to 50 % and 33 % respectively. Same datasets have been tested on different algorithms and results demonstrated that performance had been improved by applying the genetic algorithm based technique for feature selection [25].

A hybrid GA has been used that has improved the performance and also reduce the size of the subset substantially. Experimentations have been performed on different datasets and results showed that hybrid genetic algorithm has much better performance than a simple genetic algorithm. Two proposed parameters have been developed that have effective performance on hybrid genetic algorithm [28].

4. Genetic Algorithm Weaknesses in Feature Selection

It has been observed that the genetic algorithm finds abest optimal solution in local search; however, thesimple genetic algorithm has some weaknesses like premature convergence. Therefore to improve the fine tuning and efficiency of the genetic algorithm for FSS has been used in the local search.

GA algorithm for hybrid feature subset selection labeled some of the identified features that are not evaluated previously which can create problems further due to the level of information they contain in them. Longer run time is also required to select the optimized features because the proposed framework for hybrid feature selection took a lot of time to converge and canbe easily handled by training or representing least subset of selected features with less amount of time and can resolve most of the performance and time-related issues for hybrid approach [22].

One of the major issues is that they have discovered isolated feature and rank on their predictive accuracies, higher accuracy feature gets selected, that does not ensure that those selected features have no redundancy and no noise in them.There should be some mechanism to solve this issue so that proper transparency in feature selection should be maintained [21].

4.4 Differential Evolution

Differential evolution (DE) is a population-based optimization technique; the Initial population is initialized in a search space and then generates a new set of vectors from that initial population by using reproduction operator of DE. The roulette wheel selection process has been used for removing the redundant features from the original dataset. After applying this process, new population will emerge in a search space [29].

1. Current Research

Khushaba et al. [29] proposed a technique for feature subset selection in which they discussed a mixture of probabilistic technique and evolutionary technique by using differential evolution and statistical repair method for selecting optimized features. They have shown that DE is very useful to select the good features.

2. Differential Evolution in Feature Selection

Differential evolution has been used with repair mechanism to select the distributed features. Most optimized features are identified and given a high probability of selection. An optimal subset of features has been extracted by using differential evolution having different dimensionalities [31], [41], [53].

3. Differential Evolution Performance in Feature Selection

Classification accuracy has been tremendously improved by using DE. Results presented in this technique have shown significant improvement in the performance as compared to other feature subset selection techniques. Results have been generated on multidimensional datasets with various dimensionality and target classes, compared with many other population-based features. It enhanced the performance and selected the optimized features that in result build the optimum model [29].

4.5 Genetic Programming

Genetic Programming (GP) is an evolutionary hyper-heuristic search algorithm used for building dynamically building a logical model and mathematical models. The main advantage of genetic programming makes it the best algorithm for the construction of multiple features from the input space. State of the art algorithms made assumptions and constraints for building the solution of features, but genetic programming can build multiple transformations without being involved in assumptions and constraints issues. Genetic programming gives very promising results on the input set of features and constructs multiple features very effectively [47].

1. Current Research in GP

Neshstian et al. [29] proposed a methodology of using a genetic programming on a filter based method of feature selection

for construction of features for learning classifiers. Hybrid feature selection using a GP, the wrapper-based approach is used by taking a filter based method for construction of features. From a poll of input features, those highly optimal features are selected and constructed. An objective function is also proposed to construct the multiple features from input set. New constructed features used the input space to transform them into new space.

2. Genetic Programming Performance in Feature Selection

Extensive experimentation has been performed in the paper on multiple problems and classifiers. Accurate results are achieved by using rule-based and decision trees based classifiers, and the constructed features have high accuracy and performance.

3. Why evolutionary algorithms for feature selection?

Evolutionary algorithms have successfully been used in large scale applications, to handle large dimensional datasets with a huge amount of features; evolutionary algorithms perform very effectively to extract the useful features. Feature selection is a critical and highly computational complex task, due to high optimization capability they construct heuristics for complex optimization problems and solve them. The concept of the population in evolutionary algorithms makes them east to parallelize. They work in such a way, some features to exchange, which one to exchange, which one to replace, the synchronous or asynchronous exchange between their co-related features. Robustness is the major motivation to use evolutionary algorithms for feature subset selection; in this case, robustness is that each run of the algorithm it selects the same number of features and most quality features that gave high predictive accuracy. Evolutionary algorithms are used when the problem is too complex to handle, and no exact method is available for the problem [2,3,4,5].

5. Swarm Intelligence

Swarm Intelligence (SI) is another highly effective branch of soft computing and Swarm Intelligence also works on natural processes environment. Swarm optimization consists of a collection of mobile agents that collectively work to solve complex optimization problems, these agents communicate with each other directly or indirectly. Swarms emerge a collective global pattern which optimizes multi-modal and complex numerical optimization problems. The idea of SI came from natural social insect's behavior and their communication; it is a very interesting field used in distributes systems of agents, performance optimization, distributed allocation of tasks, evolutionary gaming and in indirect interactions [3], [4], [5], [6], [7], [8], [9], [10], [23], [24], [30], [35].

5.1 Swarm Intelligence in Feature Selection

Large dimensional data is difficult to handle with conventional mathematical algorithms, by using these algorithms redundancy and anomaly issues cannot be resolved [43]. We need quick optimization and selections of features from the large dimensional datasets; swarm intelligence has the capability to find out the relevant features from high-dimensional data. Feature process can be optimized by using swarm intelligence based techniques and literature showed that these SI algorithms are robust. Two most popular techniques of SI are particle swarm optimization (PSO), and AntColony optimizations (ACO) have successfully been used to extract features from large dimensional datasets [4].

5.2 Swarm Algorithms Behaviors

The swarm consists of simple individuals that use their local information to perform tasks and through certain protocol they interact with each other, in this way swarm achieves its objectives. Due to the interaction among the members of a swarm, some collective global patterns evolve referred as emergent behavior. Since individuals are semi-autonomous i.e. they don't follow any centralized command and control system, rather they act independently and interact with other members to form a coherent body of the swarm [35].

Swarm intelligence plays an important role in features subset selection. Selection of optimized features can only be possible by effective and efficient dynamic optimization algorithms that are capable of selecting a high accurate subset of features. Two main types of Swarm Intelligence i.e. Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) have been discussed below [23], [27], [35].

5.3 Particle Swarm Optimization

The particle swarm optimization algorithm is stochastic search technique that is based on the movement and intelligence behavior of swarms in the natural environment and works on the behavior of social insects. Optimal or best solution is the final destination of the particle swarm optimization. Particle in particle swarm optimization is flown through multi-dimensional search space. Each position of the particle is updated by using the personal best position and global best position of each particle. PSO was first proposed to simulate the social behavior of birds flocking [3].

5.4 The Algorithm

Swarms are composed of many particles, where each particle has its current position in the search space; each particle has its velocity totrack the speed and direction in which particle is currently moving. Similarly, each particle in the search space has its fitness value. Particle swarm optimization works on the principle of collective behavior which consists of five principles [35].

The first principle of collective behavior in PSO is homogeneity. It states that every individual in the swarm consists of some characteristics as compared to another swarm in the search space. They form a collective bird flocking system and moves without a leader, but leaders from the collective swarms can be chosen out by randomized search.

The second principle of collective behavior is a locality of each swarm. Each is an independent agent in its search space that works on the bottom up approach. Flocks movement is considered by a vision that is most important for the flock coherence.
The third principle of collective behavior is collision avoidance. Swarm always travels collectively and in parallel due to which they can collide with each other. Flock mates collision can be avoided by introducing a collision avoidance mechanism that will prohibit and stop flock from the collision.
The fourth principle of collective behavior is velocity matching. Swarms travel in search space; every swarm has a different velocity which causes disturbance of the whole swarm. It tries to develop a mechanism to equalize the velocities of flocks so that coherency should be maintained.
Fifth and the last principle of collective behavior is flock centering. In the swarms leaders emerge to lead the whole swarm, individuals in the population tend to centralize them self towards those leaders.

5.1 Particle Swarm Optimization in Feature Selection

Our focus is on feature selection using particle swarm optimization algorithm. In PSO there is a particle which iteratively evaluates the fitness of each feature and selects the best features, that particle is called as *particle best or local be st*and the best feature among all features is known as *global best.* The position of each feature will be defined accordingly [40], [44]. Accuracy has been improved by using candidate solutions of PSO, and most useful features have been extracted. To find relevance different methods have been used to get maximum performance, and PSO has been proved to be a useful optimization technique for feature selection. Filter and Wrapper based methods have been used with the combination of particle swarm optimization to improve the performance and predictive accuracy of selected features from the high dimensional dataset. By using PSO process of feature selection has been improved and it performed well on large dimensional data as well [27].

5.2 Current Research and Future Directions

Wang et al. [22] proposed a technique for feature subset selection based on rough sets theory and particle swarm optimization, previously rough sets had been recommended as thebest technique to find the subset of features because it has high success rate. It has some limitations that lead to predicting misleading features that result in a decline in predictive accuracy as well as optimality cannot be perfectly guaranteed of the model. Datasets are very large it is very difficult to search complete data fully because the high amount of computations are required, so to cater this situation researchers have proposed stochastic and probabilistic techniques for optimal feature subset selection. In particle swarm, optimization there exists swarms of particles having a certain velocity, and they are flying in their search spaces [44,45].The major purpose of those swarms is to find optimal regions in very complicated search space by communicating with the particular individuals in the search space. PSO has discovered the optimal feature to improve efficiency and performance of the learning model and it also to improve the predictive accuracy. By comparing two evolutionary algorithms like particle swarm optimization and genetic algorithm, PSO is not much complex as genetic algorithm because it has no complex operators of crossover and mutation [22], [23], [56], [60].

5.3 Particle Swarm Optimization Strengths in Feature Selection

PSO is computationally inexpensive and cheaper than a genetic algorithm; it requires less memory and time. The experimentation results on different datasets had demonstrated that PSO has extracted good features as compared to GA and rough sets, and it also perform efficiently [26], [36].

5.4 Particle Swarm Optimization Weaknesses in Feature Selection

There are some of the improvements that can be incorporated like they have not used probabilistic approach for finding best and optimal features. If the hybrid algorithm is used instead of the single algorithm, thenperformance can be enhanced.The position of particle and velocity are major components in particle swarm optimization by optimizing their value we can enhance its performance [53].
Particle swarm optimization and rough set used for selecting subsets of features have some of the performance drawbacks like PSO uses trial and error method to select maximum velocity. In feature subset selection velocity has to be in limit up to N, since 1 to N is the dynamic range for selecting the subset of the feature, but the major disadvantage of this approach is that each particle will have a poor exploration of finding the optimal features. Larger inertia weights will help to find the optimal amount of features. Extension in this research can be to introduce a hybrid algorithm; fitness function and position of the particle are the key factors for feature subset selection, which can increase performance and predictive accuracy. Parallel algorithms can also be helpful in increasing performance [36].

6. Ant Colony Optimization

Ant colony optimization is another bio-inspired natural population-based meta-heuristic algorithm which is proposed by Marco Dorigo[6], [11]. Ant colony optimization is based on food foraging behavior of ants, Thisnature-inspired algorithm works on optimization done by ants in their environment. Ants have a remarkable capability of memorization to learn from their environment and to collaborate with each other. This marvelous behavior of ants exhibits in this algorithm. Ant colony

optimization algorithm has been used in various applications. It works similar to the principle of particle swarm optimization. Swarms cooperate with themselves similar to that; ants also coordinate with each other to form collective behavior in an environment this interaction process in ant colony optimization is called as stigmergy. In the search space of ants, there are individuals that coordinate this principle is based on coordination between ants. The basic purpose of ant colony optimization algorithm is optimization of paths to represent a complete solution in search space that is allocated to it. Working of Ant algorithm is composed of many processes that are Ants deposits pheromone on their path for the help of successor Ants to travel on the path where the pheromone is deposited. If the path has a high amount of pheromone deposited than Ants will choose that path than the path where less amount of pheromone is deposited. That path is the optimized path where the objective is placed [27].

6.1 The Algorithm

Ant colony optimization is a probabilistic search approach based on the pheromone and heuristic, which tells the information about the quality of information. The pheromone values are updated along with the evaporation of pheromone. The optimal path will be that where dense pheromone will be deposited and that route will be highly optimal. Ant colony optimization has some major components that are discussed below [6,43,55,58].

First is the representation of search space which usually represented in the form of a graph.

The secondcomponent is the heuristic that tell the quality of a solution component.

Third is the feedback process in which old pheromone that is deposited by ants is updated according to the quality of the solution.

The fourth is the constraint satisfaction mechanism in which it evaluates the achievability and feasibility of solution constructed?

Last on is the solution construction method, in which algorithm takes the parameters and produces the candidate solutions and calculate the state transition probabilities.

6.2 Ant Colony Optimization in Feature Selection

Ant colony optimization has been extensively used in combinatorial optimization and discrete optimization problems. Large dimensional datasets are used for this purpose to solve NP-hard problems [33]. Most of the problems related to feature subset selection are also NP-hard problems, which require combinatorial optimization using ACO. It finds useful features from the dataset and maps that selected subset of features in objective function. Ant colony optimization has been widely used in data mining application like classification, clustering, and feature selection. Ant colony optimization is usedin feature subset selection in two different ways, i.e. different variants of ACO are employed to feature selection techniques. Subsets are evaluated bysome statistical measures. Population-based feature selection is very commonly used the technique; featuresare selected on the amount of the probability that each feature contains, high probability feature with useful information in it will be selected on the first basis, it is discrete optimization problem [29]. Ant colony optimization increased the search speed by checking the maximum features where the maximum amount of pheromone is deposited and selects subsets from these features. This technique reduced the computational complexity which will result in the increase in the performance and predictive accuracy of the algorithm [6].

6.3 Current Research and Future Directions

Sivagaminathan et al. [23] proposed a technique for feature subset selection in which they have used a hybrid technique for feature subset selection using neural network and ant colony optimization to select the subset of features. They have used accuracy as a goodness measure to rank the features. In classification and analysis of the model, they have used artificial neural networks and ant colony optimization and also used a mixture of filter based technique and wrapper based technique. As a combination, they have used hybrid feature subset selection and applied it on datasets of medical diseases.

This methodology proposed a hybrid approach using neural networks and ant colony optimization; its structure is composed of following steps.

Hybrid Approach with ANN and ACO: This method is like a chemical reaction between two methodologies that is an indirect communication mode takes place which is called as stigmergy. ACO is a MetaheuristicAlgorithm in which pheromone plays very important role for distance to the food source. Where larger the value of pheromone deposited that path will have more the chance to be selected. In this approach, artificial neural networks have been used as a classification function.

Hybrid ANN: There are N numbers of features in the dataset and n increases at a constant rate. The dataset is composed of training data and test data. The neural network will train until 30 epochs.

Ant Colony Algorithm: We have discussed ACO algorithm in detail, in this step there will be some of the steps that are followed by the ant colony optimization algorithm. Those are state transition rule, global updating rule, and local updating rule.

Al-Ani et al. [26] proposed a technique for feature subset selection based on ant colony optimization, which finds and search optimized route from their nest to food source, that path is the shortest and optimized path in search of food that's why it is called as ant colony optimization. Speech segments dataset is used for the classification of features, and efficient results are generated by using this technique. This proposed technique improved the performance of feature subset selection and achieved higher accuracy rates when compared with other techniques.

Kabir et al. proposed a technique for feature subset selection in which they have used hybrid ant colony optimization algorithm, whose purpose is to enhance the performance and predictive accuracy by selecting the most relevant feature subsets [32]. This technique selects the features that are less in size and complexity. This algorithm uses hybrid search technique that contains both wrapper and filter based approach. Ant colony optimization has two main features which are responsible for its running that is pheromone which is deposited by ants in their search path and the amount of heuristic information that is associated with each pheromone. A new set of rules that contain both pheromone value and heuristic information will be designedto facilitate the selection of a subset of features.
Every step of algorithm contains help for an ant to guide their path for searching and set of new rules have been constructed for guiding the ant in the right search direction, also searching graphs will also help ants in the right direction to search. These combinations of graph and rules will provide ants to help in their exploration of new regions to search and get right benefits from what users require from the dataset. This new technique of hybrid ant colony optimization is very beneficial for searching optimum global features with best performance and accuracy. Different types of datasets with different dimensionalities have been tested on ACO search algorithm, in which majority of them are classification datasets with 10-2000 dimensions in them. This technique is also tested on many existing feature selection algorithms, and the comparison result shows the remarkable performance of feature selection using ant colony optimization as compared to other search techniques. Reduced subsets of features have been generated by using this algorithm [26,32].
Ali et al. [33] proposed a technique for feature subset selection in which they have discussed two techniques; oneis based on statistical analysis, and other is ant colony optimization. Filter based approach is used for selecting a subset of a feature to improve predictive accuracy and performance.

6.4 Ant Colony Optimization Strengths in Feature Selection

Feature subset selection is not only concerned with eliminating or reducing the required features but also with the features that are noisy, this method help to select optimal features from such dataset so that predictive performance and accuracy can be improved. There exists a relationship between the features present in the dataset. Using the approach of neural networks and ant colony optimization, it can be seen that they have shown best predictive results by selecting the optimal subset of features. This work can be improved to develop heuristic models for medical diagnostics application and will help feature selection procedure to select that feature where the risk, cost, and high diagnostic value is associated. Furthermore, comparisons that are discussed in this paper have some of the applications that can be enhanced like pheromone updating criteria for ANT algorithm to strengthen the parameter in algorithm [23].
Metaheuristic algorithms perform very effectivelyin selecting the most optimized features from the dataset. Like genetic algorithm and proposed ant colony optimization outperformed despite there is a large dimensional dataset. Ant colony optimization uses its updating of pheromone value and frequency of ants. This technique works perfectly on a small number of features but as the features increase it shows that ACO performance is better than other Metaheuristic algorithms. Some of the improvements in internal parameters of ANT algorithm like pheromone updating, rule quality and heuristics can be made to achieve the best performance to select an optimal subset of features from the dataset [26], [60], [62].
The idea behind ant colony optimization for feature selection is very clear that is guiding the ants using a hybrid technique to search. Information gain measurement procedure has been introduced in this method whose major advantage is computationally less expensive. Following are the section for ACOFS. Determination of subset size, subset evaluation, best subset selection and hybrid search process are the main steps that are involved in ACOFS. This algorithm performs very well on large datasets as it has very less computational complex than a simple ant colony algorithm [32].
Symmetric uncertainty is a probabilistic method that has been used with a MetaheuristicAlgorithm ant colony optimization it combines to develop a method for selecting optimal subsets of features from the datasets. It results in improved predictive performance and accuracy of the algorithm. Certain conditions used in both ACO and symmetric uncertainty can be enhanced to improve the performance of feature selection [36]. The main purpose for using filter based technique is that it used less computational efficiency and it is highly efficient in term of its predictive accuracy. It works well on less number of features very effectively, feature value and threshold value is selected in most of the cases. It selects features from the dataset by their worth. A Large number of experimentations has been performed with three best and known classification algorithms in this regard to get good results. Results have shown best performance accuracy [33], [54], [58], [59].

6.5 Ant Colony Optimization Weaknesses in Feature Selection

Metaheuristic technique ACO that has been used for feature subset selection may fail to perform under some of the conditions. Parameters of ant colony algorithm like pheromone updating, rule quality and heuristic can be improved based on problems assumptions so that performance and predictive accuracy for selecting optimized features can be improved and a best possible subset of features can be extracted [32].

6.6 Why swarm intelligence for feature selection?

Swarm intelligence is a revolution in soft computing it consists of highly effective metaheuristic algorithms used to solve high optimization problems and data mining problems. Two main Metaheuristic algorithms are particle swarm optimization and ant colony optimization. The most important argument to use these algorithms is that they are less complex even from evolutionary algorithms. Feature selection can be given a new life, due to their optimum parameters functions, their

convergence and robustness [3], [4], [5].

7. Bio-Inspired Algorithms

7.1 Current Research

Yun et al. [27] proposed a technique for feature subset selection in which they have discussed biologically inspired algorithms for extracting effective features, these biological techniques are very effective and have remarkable performance than other conventional techniques of data mining. Full dataset searching is very problematic and computationally expensive task, finding optimal feature is a big challenge to improve accuracy and predictive performance. In this proposed technique they have used biological algorithms that are related to real life such as genetic algorithm and particle swarm optimization. The datasetcontains relevant as well as irrelevant features; those irrelevant features are not helpful in any perspective they only increase the computational complexity of our model.

7.2 Bio-Inspired Algorithms in Feature Selection

Variations of these approaches have been tested and applied to feature subset selection on various real-time datasets they have also tested on hybrid evolutionary technique by combining both genetic algorithm and particle swarm optimization and performance improve. Algorithm results show that the performance of relevant genetic algorithm and relevant particle swarm optimization is better than single GA and PSO for feature subset selection in regards to both accuracies of learning of model and speed of evolution. Real life datasets are tested, and results showed better performance of PSO [29].

7.3 Performance of Bio-Inspired Algorithms on Feature Selection

The performance of RMR in this technique is very effective in term of getting relevant information from the dataset, which results in a deduction in redundancy and improves predictive accuracy. If a solution of GA and PSO is merged with ACO learning accuracy and predictive performance can be enhanced. Hybrid techniques can also be developed which can enhance the performance of feature selection procedure, but they will take longer time in evolving their strategies of selecting optimized features. These hybrid techniques can be developed by combing ACO with a filter or wrapper-based approach of feature subset selection. The hybrid approach works very well in all bio-inspired algorithms and very useful in finding optimal solutions. Features that are produced using these bio-inspired algorithm has different complexities; our purpose is to enhance the predictive accuracy and performance of the algorithm by analyzing the performance of every approach on different datasets [27], [39].

8. Feature Selection by Classification, Learning Techniques, and Large Dimensionality Domains

8.1 Current Research on Feature Selection using Classification

Bermejo et al. [29] proposed a technique for feature subset selection whose purpose is to gain efficient performance by making a compact and clear model of classification, for practical purposes they have used high dimensional datasets for predicting a large number of attributes from the dataset. They have not used standard wrapper algorithm as its complexity is much higher than filter based algorithm. They have proposed GRASP algorithm whose main purpose is to speed up the feature subset selection process, in less amount of time and complexity we have to produce better features so that we can perform iteration on our optimized dataset. GRASP is amulti-objective method that runs in parallel.Several experiments have been performed based on this algorithm on the dataset, after experimentations generated results have been compared with other previously existing algorithms with thehigh multi-dimensional dataset, results shows a lot of significant and viable improvement in the performance and quality of asubset of feature that has been selected.

8.2 Performance Strengths and Weaknesses of Classification on Feature Selection

Performance and predictive accuracy can only be effectively tested on high dimensional datasets. Classification is one of the best methods in data mining to evaluate high-dimensional datasets. It cooperates with the dominated and non-dominated solutions by creating a common pool with the variable they contain. Performance is the major issue that should be upgraded by using classification we have raised performance by using wrapper and filter methods of feature subset selection. Metaheuristic algorithms can be applied to feature subset selection so we can test cluster of features not a single solution to obtain better predictive accuracy [17].

8.3 Current Research on Feature Selection using Learning Techniques

Liu et al. [18] proposed a technique for feature subset selection that is based on two techniques of data mining, one is supervised learning a technique that is classification, and other is unsupervised learning technique that is clustering. It compares different old techniques of classification and clustering with a new one based on different search strategies, evaluation methods and the task of data mining; theyhelp and provide proper guidelines for selecting the best feature from the dataset. They have set a framework that establishes a system to join different feature selection. The combination of more than two algorithms can add massive advantage over only running a single algorithm they have shown this through examples and experimentations by merging meta-algorithms that show effective performance than a single one. In data mining, combined algorithms are adding diverse effects than using isolated algorithms. But there can occur some of the major and minor issues related to complexity and performance of merged algorithms.They have not proposed a specific or generalized solution for combined algorithm's complexity and performance. Two or more combination of algorithms when run on huge dataset they will have collective complexities which will be obviously greater than a single algorithm. In feature subset selection we have analyzed thecomplexity of algorithm because it is the major concern for us if it is high

than it is adding drastic negative effects on performance and speed of prediction accuracy since our purpose is to increase predictive accuracy by reducing the complexity of algorithm [27].

8.4 Performance Strengths and Weaknesses of Learning Techniques on Feature Selection

The basic objective of integrating feature selection for classification and clustering is to introduce the concept and understandings of feature selection, to make unified platform in which we can easily handle data mining tasks. Curse of dimensionality is the major issue in feature selection. Classical algorithms for feature selection have quadratic, or higher complexity about N. filter based method is computationally less expensive than wrapper but now for large dimensional data, ahybrid approach is quite feasible. These algorithms focus on combining both to achieve the best performance and accuracy they usually have less complexity. In N-dimensionaldataset numbers of instances are also I so this needs extensive, efficient approach. There are some of the challenges by applying this technique to feature selection; the preprocessing step can take a lot of time that require multiple data scans or random access to data which will be very costly. Clustering cannot cover all dimensions there can be some loopholes which will create problems in feature selection in the end performance issue will arise [18].

8.5 Current Research on Feature Selection using Large Dimensionality Domains

Gheyas et al. [31] proposed a technique for feature subset selection in which they have discussed large dimensions and large datasets for selecting optimum feature subsets in them. It is very difficult task to search complete dataset; exhaustive searching is impossible, and it will be an NP-complete problem. To cater this problem they have developed a hybrid search algorithm that can keep away from trapping into the local minima. In a genetic algorithm, crossover operator rate of convergence is very high due to its high computational search ability and ability to extract optimized features is the type of generalized regression technique of neural networks.

8.6 Performance Strengths and Weaknesses of Large Dimensionality Domains on Feature Selection

The performance of this technique has been checked on various datasets that contain a large number of dimensionalities and compared with other well-knowndatasets; it shows better results and output as compared to other search algorithms for selecting best and optimized feature that are helpful in improving the predictive performance of feature selection model. This technique solves the problems of feature subset selection without involving filter based model. It's the combination of search algorithms SAGA which is a highly effective and optimized algorithm. It does not trap in local minima; it's a hybrid genetic algorithm both crossover and mutations operators strengthen overall performance of the algorithm. SAGA is composed of following qualities.

Global search algorithms allow maximum search in our space to find optimum features without getting stuck in local minima with the high computational capability and high rate of selecting a subset of features. This new technique is based on mutation concept of genetic algorithm. This algorithm is tested on all evolutionary algorithms and performance results are checked. Feature subsets are selected from the large dimensional dataset, and computational efficiency matters a lot in large dimensional data. Therefore, this technique provides best and optimal results in feature subset selection approach [31].

9. Conclusion and Future Works

In this paper, we provide the review and analysis of the state of the art techniques of soft computing applied for the feature subset selection. We have discussed in detail the strength and weaknesses of each technique. We have focused on meta-heuristic, population-based and bio-inspired algorithms that are used in solving the problems of features selection. We have given a detail description of these algorithms, their strengths and weaknesses and performance in feature selection. We have discussed the research trends of soft computing algorithms in feature subset selection and for this purpose, we have discussed different techniques and analyze them critically.

Knowledge discovery is an ultimate task of data mining, and useful knowledge is very precious, features selection is an emerging and diverse area of research. Data is increasing day by day and need an effective process of mining and also to discover the useful features in it. Performance and classification accuracy is a big challenge and researchers are doing hard in finding the solutions to cater it. The main advantage of feature subset selection is three fold, it enhances the potential of classifiers by opting for useful predictors, it is cost effective and fast, and it provides an understanding of the underlying processes that generated the data. The main objective of feature selection is to opt for those subsets of features from the original feature space that provide useful information. A useful feature is neither irrelevant nor redundant.

There are two main categories of feature selection discussed filter based and wrapper based methods. Filter based approach is statistical approach and has less computational complexity at the costof compromising on the predictive accuracy.A wrapper-based approach is an inductive approach to feature selection with high computational cost and high predictive accuracy.

Acknowledgement

The author would like to thanks Dr. Waseem Shahzad Professor NUCES, Islamabad for his assistance and corporation. Also special thanks to HEC digital library for providing research material.

REFERENCES

[1] J. Han, and M. Kamber, Data Mining: Concepts and Techniques, 2nd Edition, *Morgan Kaufmann Publishers*, 2006.

[2] W. J. Frawley, G. P. Shapiro, and Christopher J.," Knowledge Discovery in Databases: An Overview", *AI Magazine*, Vol. 13, No. 3, pp. 57-70, 1992.

[3] A.P. EngelBrecht, "Evolutionary Computations & Swarm Intelligence", in *Computational Intelligence an Introduction,* London, UK: John Wily & Sons Inc., 2002

[4] K.A. De Jong, "Introduction", in *Evolutionary Computations a Unified Approach,* London, UK: MIT Press, 2006

[5] A. Konar, *Computational Intelligence Principles, Techniques and Applications.* Springer-Verlag 2005.

[6] A.P. Engelbrecht, *Fundamentals of Computational Swarm Intelligence.* John Wiley & Sons, 2005.

[7] R.Jensen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches, Wiley*-IEEE Press 2008.

[8] M. Dorigo, and T. Stützle, *Ant Colony Optimization.* Cambridge, MA: MIT Press, 2004.

[9] A.E. Eiben, and J.E. Smith, *Introduction to Evolutionary Computing.* Natural Computing Series, 2nd edition, 2007.

[10] D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, 1989.

[11] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant System: Optimization by a colony of cooperating Agents," IEEE Transactions on Systems, Man, and Cybernetics, Part B, Vol. 26, No. 1, Feb. 1996.

[12] Q. Song, J. Ni and G. Wang, "A Fast Clustering-Based Feature Subset Clustering Algorithm for High-Dimensional Data," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 1, pp. 1–14, 2013.

[13] F. Tan, X. Fu, Y. Zhang and A.G. Bourgeois, "Improving Feature Subset Selection Using a Genetic Algorithm for Microarray Gene Expression Data," Proc. IEEE Congress on Evolutionary Computations, Vancouver, BC, Canada, pp. 2529–2534, 2006.

[14] P. Mitra, C.A. Murthy and S.K. Pal, "Unsupervised Feature Selection Using Feature Similarity," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 301–312, 2002.

[15] H.L. Wei and S.A. Billings, "Feature subset selection and ranking for data dimensionality reduction," IEEE Transactions on Pattern Analysis and MachineIntelligence, Vol. 29, No.1, Jan. 2007.

[16] Z. Zhao, L. Wang, H. Liu and J. Ye, "On Similarity Preserving Feature Selection," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 3, pp. 619–632, 2011.

[17] P. Bermejo, J.A. Gamez and J.M. Puerta, "A GRASP Algorithm for Fast Hybrid (Filter-Wrapper) Feature Subset Selection in High Dimensional Datasets," Pattern Recognition Letters, vol. 32, no. 5, pp. 701–711, 2011.

[18] H. Liu and L. Yu, "Towards Integrating Feature Selection Algorithm for Classification and Clustering," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 4, pp. 491–502, 2005.

[19] W. Fan, N. Bouguila and D. Ziou, "Unsupervised Hybrid Feature Extraction Selection for High Dimensional Non-Gaussian Data Clustering with Variational Inference," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 99, pp. 1–18, 2012.

[20] J. Huang, Y. Cai and X. Xu, "A Hybrid Genetic Algorithm for Feature Selection Wrapper Based on Mutual Information," Pattern Recognition Letters, vol. 28, no. 13, pp. 1825–1844, 2007.

[21] F. Tang, X. Fu, Y. Zhang and A.G. Bourgeois, "A Genetic Algorithm Based Method for Feature Subset Selection," Soft Computing- A Fusion of Foundations, Methodologies and Applications, vol. 12, no. 2, pp. 111–120, 2007.

[22] X. Wang, J. Yang, X. Teng, W. Xia and R. Jensen, "Feature Selection Based on Rough Sets and Particle Swarm Optimization," Pattern Recognition Letters, vol. 28, no. 4, pp. 459–471, 2007.

[23] R.K. Sivagaminathan, and S. Ramakrishnan, "A Hybrid Approach for Feature Subset Selection Using Neural Networks and Ant Colony Optimization," Expert System with Applications: An International Journal, vol. 33, no. 1, pp. 49–60, 2007.

[24] R.S. Parpinelli, H.S. Lopes, and A.A. Freitas, "Data mining with an ant colony optimization algorithm," IEEE Transactions on Evolutionary Computation, Vol. 6, No. 4, pp. 321–332, Aug. 2002.

[25] M.E. ElAlami, "A Filter Model for Feature Subset Selection Based on Genetic Algorithm," Knowledge Based Systems, vol. 22, no. 5, pp. 356–362, 2009.

[26] A. Al-Ani, "Ant Colony Optimization for Feature Subset Selection," World Academy of Science, Engineering and Technology, vol. 4, no. 4, pp. 529–532, 2007.

[27] C. Yun, B. Oh, J. Yang and J. Nang, "Feature Subset Selection Based on Bio-Inspired Algorithms," Journal of Information Science and Engineering, vol. 27, no. 5, pp. 1667–1686, 2011.

[28] II.S. Oh, J.S. Lee and B.R. Moon, "Hybrid Genetic Algorithm for Feature Selection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1424–1437, 2004.

[29] R.N. Khushaba, A. Al-Ani and A. Al-Jumaili, "Feature Subset Selection Using Differential Evolution and a Statistical Repair Method," Experts Systems with Applications: An International Journal, vol. 38, no. 9, pp. 11515–11526, 2011.

[30] A. R. Baig, W. Shahzad, and S. Khan: Correlation as a Heuristic for Accurate and Comprehensible Ant Colony Optimization Based Classifiers. IEEE Transactions on Evolutionary Computation 17(5): 686-704 Oct, 2013

[31] I.A. Gheyas and L.S. Smith, "Feature Subset Selection in Large Dimensionality Domains," Pattern Recognition, vol. 43, no. 1, pp. 5–13, 2010.

[32] M.M. Kabir, M. Shahjahan and K. Murase, "A New Hybrid Ant Colony Optimization Algorithm for Feature Selection," Experts Systems with Applications: An International Journal, vol. 39, no. 3, pp. 3747–3763, 2012.
[33] S.I. Ali and W. Shahzad, "A Feature Subset Selection Method Based on Symmetric Uncertainty and Ant Colony Optimization", "International Journal of Computer Applications (0975-8887), vol. 60, no. 11, pp. 5–10, 2012.
[34] W. Shahzad, and A.R. Baig, "A hybrid associative classification algorithm using ant colony optimization," International Journal of Innovative Computing, Information and Control (IJICIC), 2010.
[35] C. Grosan, A. Abraham and M. Chis, "Swarm Intelligence in Data Mining," Studies in Computational Intelligence, vol. 34, no. 3, pp. 1-21, 2006
[36] Y. Liu, G. Wang, H. Chen, H. Dong, X. Zhu and S. Wang, "An Improved Particle Swarm Optimization for Feature Selection," Journal of Bionic Engineering: An International Journal, vol. 8, no. 2, pp. 191–200, 2011.
[37] J. Handl and J. Knowles, "Feature Subset Selection in Unsupervised Learning Via Multi-objective optimization," International Journal of Computational Intelligence Research, vol. 2, no. 3, pp. 217–238, 2006.
[38] C.L. Huang and C.J. Wang, "A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines," Experts Systems with Applications: An International Journal, vol. 31, no. 2, pp. 231–240, 2006.
[39] F.G. Lopez, M.G. Torres, B.M. Batista, J.A.M. Perez and J.M.M. Vega, "Solving Feature Selection Problem," European Journal of Operational Research, vol. 169, no. 3, pp. 477–489, 2006.
[40] K. Sarojini, K. Thangavel and D. Devakumari, "Supervised Feature Subset Selection Based on Modified Fuzzy Relative Information Measure for Classification Chart," International Journal of Engineering Science and Technology, vol. 2, no. 5, pp.2456–2465, 2010.
[41] L. Zhou, L. Wang, and C. Shen, "Feature selection with redundancy constrained class separability" IEEE Transactions on Neural Networks, vol. 21, no.2, pp.853–858, 2010.
[42] M. Dong and R. Kothari, "Feature Subset Selection using a new definition of classifiability," Pattern Recognition Letters, vol. 24, no. 3, pp. 1215–1225, 2003.
[43] A.G. Karegowda, M.A. Jayaram and A.S. Manjunath, "Feature Subset Selection Using Cascaded GA & CFS: A Filter Based Approach in Supervised Learning," International Journal of Computer Applications, vol. 23, no. 2, pp. 0975–0984, 2011.
[44] A. Suebsing and N. Hiransakolwong, "A Novel Technique for Feature Subset Selection Based on Cosine Similarity," Applied Mathematical Sciences, vol. 6, no. 133, pp. 6627–6655, 2012.
[45] E.C.C. Tsang, D.S. Yeung and X.Z. Wang, "Optimal Fuzzy-Value Feature Selection Problem," IEEE Transactions on Fuzzy Systems, vol. 11, no.2, pp. 202–213, 2003.
[46] M.L. Raymer, W.F. Punch, E.D. Goodman, L.A. Kuhn and A.K. Jain, "Dimensionality Reduction Using Genetic Algorithm," IEEE Transactions on Evolutionary Computations, vol. 4, no. 2, pp. 164–171, 2000.
[47] K.Neshatian, M.Zhang and P.Andreae, "A filter Approach to Multiple Feature Construction for Symbolic Learning Classifiers Using Genetic Programming," IEEE Transactions on Evolutionary Computations, vol. 16, no. 5, pp. 645–661, 2012.
[48] W. Shahzad, and A.R. Baig, "A Hybrid Associative Classification Algorithm using Ant Colony Optimization", International Journal of Innovative Computing, Information and Control (IJICIC), 2010.
[49] C. Tsai, C. Tsai, H. Wu, and T. Yang, "ACODF: a novel data clustering approach for data mining in large databases", presented at Journal of Systems and Software, Vol. 23, No. 1, pp.133-145, 2004.
[50] M. Dash and H. Liu, "Consistency-based search in feature selection", presented at Artificial Intelligence, Vol. 151, No. 1-2, pp.155-176, 2003.
[51] F. Fleuret, and I. Guyon, "Fast Binary Feature Selection with Conditional Mutual Information", in Journal of Machine Learning Research, Vol. 5, pp. 1531-1555, 2004.
[52] Neshatian, M. Zhang, and M. Johnston, "Feature construction and dimension reduction using genetic programming," in Proc. Adv. Artif.Intell., LNCS 4830, 2007, pp. 160–170.
[53] J. Zhou, R. Ng, and X. Li, "Ant colony optimization and mutual information hybrid algorithms for feature subset selection in equipment fault diagnosis", in Proceeding of 10th International Conference on Control, Automation, Robotics and Vision, pp.898-903, 2008.
[54] C. Zhang, and H. Hu, "Feature selection using the hybrid of ant colony optimization and mutual information for the forecaster", in Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Vol.3, pp. 1728–1732, 2005.
[55] R. Jensen and Q. Shen, "Fuzzy-rough data reduction with ant colony optimization", presented at Fuzzy Sets and Systems, Vol. 149, pp.5-20, 2005.
[56] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization", presented at Pattern Recognition Letters, pp.459-471, 2007.
[57] M. F. Zaiyadi and B. Baharudin, "A Proposed Hybrid Approach for Feature Selection in Text Document Categorization," in World Academy of Science, Engineering and Technology, Vol. 72, pp. 137-141, 2010.
[58] R. K. Sivagaminathan and Sreeram Ramakrishnan, "A hybrid approach for feature subset selection using neural networks and ant colony optimization", International Journal of Expert Systems with Applications, Vol. 33,No. 1, July, 2007.

[59] S. Nemati, M.E. Basiri, N. Ghasem-Aghaee, and M.H. Aghdam, "A novel ACO-GA hybrid algorithm for feature selection in protein function prediction", presented at Expert Syst. Appl., Vol. 36, No. 10, pp.12086-12094, 2009.
[60] H. Gao, H. Yang, and X. Wang, "Ant colony optimization based network intrusion feature selection and detection", in Proceedings of International Conference on Machine Learning and Cybernetics, Vol. 6, pp. 3871–3875, 2005.
[61] M. E. Basiri, N. Ghasem, and M. H. Aghdam, "Using Ant Colony Optimization-Based Selected Features for Predicting Post-synaptic Activity in Proteins", EvoBIO, pp 12-23, 2008.
[62] K. R. Robbins, W. Zhang, J. K. Bertrand and R. Rekaya, "The ant colony algorithm for feature selection in high-dimension gene expression data for disease classification", in Math Med Biol, Vol. 24, No. 4, pp 413-426, 2007.
[63] Y. Marinakis, M. Marinaki, M. Doumpos, and C. Zopounidis, "Ant colony and particle swarm optimization for financial classification problems", presented at Expert Syst. Appl., pp.10604-10611, 2009.
[64] K. Krawiec, “Genetic programming-based construction of features for machine learning and knowledge discovery tasks,” Genet. Programming Evolvable Mach., vol. 3, no. 4, pp. 329–343, 2002.
[65] H. Liu and R. Setiono, “A probabilistic approach to feature selection - A filter solution", the 13th International Conference on Machine Learning, pp. 319-327, 1996.
[66] M. A. Hall, Correlation-based Feature Subset Selection for Machine Learning, PhD dissertation, 1999, Department of Computer Science.