ABSTRACT

In communication with the world, one of the important goals for human beings is to recognize objects around them. Researchers and scientists, in the area of machine intelligence, aims to design solution to real life problems in a manner analogous to human beings. Pattern recognition is an integral part of most machine intelligent systems, built for decision making.

Most of the information that surrounds mankind manifests itself in the form of patterns. Pattern recognition, naturally, is based on patterns. A pattern can be as basic as a set of measurements and attributes are any extractable measurements from it. Thus, pattern recognition aims to classify patterns based on a set of attributes obtained from measurement space.

Generally, a corpus of attributes is extracted to better characterize a pattern recognition problem. It is also important that the extracted attributes yield high predictive performance pertinent to the pattern recognition problem in hand. Attribute selection is a pre-processing tool to classification. It reduces the dimensionality of attribute spaceby selecting highly predictive attributesfrom the problem space. In a broader perspective, this stage greatly influences the recognition process, in terms of both predictive performance and in the design of a computationally simple classifier. The process of attribute selection involves two steps:

- Attribute subset search
- Subset evaluation

Wrapper based attribute selection, one of its type, employs a classifier to evaluate the attribute subset obtained during the process of search in the problem space. Support Vector Machine (SVM) is employed in this research work to evaluate the predictive performance of attribute subsets chosen during the process of attribute selection.

Researchers have explored the use of metaheuristic randomized algorithms to find an optimal attribute subset from the problem space than an exhaustive evaluation of all possible combination of attributes. Particle Swarm Optimization (PSO), an efficient search technique, has been extensively explored by researchers for attribute selection. Similarly, Flower Pollination Algorithm (FPA), with its global walk capability explores the problem space efficiently in search of an optimal solution. Both PSO and FPA are employed in this research work as search technique to implement the wrapper approach for attribute selection.

Problem of attribute selection demands a Hamming search space with every possible combination of attributes as points in it. Hence, Binary Particle Swarm Optimization (BPSO) and Binary Flower Pollination Algorithm (BFPA) best addresses the problem of attribute selection in a Hamming space than PSO and FPA respectively.

Both the optimization algorithms are population based and the individuals in the population traverses the problem space in search of an optimal attribute subset. Owing to Hamming space, there is no problem of divergence. Nevertheless, if the leader of the population (individual with best solution vector) doesnot improve itself, during the process of search for an optimal attribute subset, problem of pre-mature convergence occurs, resulting in a sub-optimal solution.

In such situation, a re-initialization strategy is necessary to restructure the leader and re-direct the individuals in the population to another position in the problem space to continue the search there on. From literature, it is understood that, researchers have solved the problem of pre-mature convergence by re-directing the leader to a random position in the problem space. In this thesis, heuristic based re-initialization strategies are proposed to handle the problem of pre-mature convergence.

In BPSO, if the leader of the population remains unchanged for a pre-fixed number of runs, re-initialization strategy, named Greedy Reset, re-directs the leader of the population to another position in the problem space with one attribute less compared to the recently stuck local optimum solution. Pre-mature convergence to a sub-optimal point in the problem space also occurs, when the velocity of individuals in the population stagnates. Localized random mistakes are introduced to overcome the problem of velocity stagnation in BPSO. BPSO with Greedy reset and localized random mistakes is used as the subset identifier and SVM as the subset evaluator to implement wrapper based attribute selection methodology.

In BFPA, re-initialization mechanism named Greedy crossover is introduced to find a new position in the problem space on pre-mature convergence to a sub-optimal solution in the problem space. Lack of explicit crossover in BFPA is the drive to design Greedy crossover on stagnation of the leader in the problem space. BFPA with Greedy crossover and SVM are engaged to design the wrapper based attribute selection methodology.

Owing to partial optimization in BPSO, re-initialization strategy named Greedy leap, is designed such that, the stagnated leader of the population is reset to another point in the problem space using the concept of flower pollination. Levy's walk on global pollination between the leader and the individuals in the population is the underlying principle of Greedy leap

mechanism. BPSO with metaheuristic inspired Greedy Leap strategy and SVM are employed to develop the wrapper approach for attribute selection.

Both BPSO and BFPA are population based metaheuristic search techniques. Population initialization is the first step in the search process. It is an important task that aids the individuals in the population to cover the entire problem space and begin the search process from all the directions. A One to all random population initialization technique is devised such that the individuals of the population represent every dimension of the problem space.

Wrapper approach for attribute selection using search algorithms with proposed re-initialization techniques and One to all random population initialization is implemented in Matlab 2013a and tested with benchmark datasets from University of California Irvine (UCI) repository representing machine learning problems.

Improvement in predictive accuracy on applying attribute selection and reduction in dimensionality of attribute space is calculated as the performance indices to evaluate the success of proposed algorithms. Performance of proposed methodology is compared with existing literature on attribute selection using BPSO and BFPA.

Also, attribute selection techniques using other metaheuristic algorithms as search strategy is reviewed and compared with the proposed methodologies. Finally, attributes selected using proposed methodologies is fed as input to six traditional classifiers such as SVM, *k* Nearest Neighbour (*k*NN), Decision Tree (DT), Discriminant Analysis (DA), Naïve Bayes (NB) and Ensemble learning. Predictive performance of classifiers with selected attributes is analyzed to validate the generality of proposed methodologies as a pre-processing technique to classification.

Finally, the three proposed wrapper approaches for attribute selection along with One to all random population initialization is applied to a pattern recognition problem in the field of agriculture. An engineering approach in solving the problem of disease identification in rice leaf is drawn and the results are analyzed. Disease spots (Brown spot and Narrow brown spot) are segmented from the rice leaf to extract colour and shape attributes characterising it. As a pre-processing step, highly predictive attributes selected by the proposed wrapper approaches is fed as input to the classifier for disease recognition.