Genetic Algorithm Paper Review

The paper, *Genetic Algorithm-Based Feature Selection*, looks into the usage of the genetic algorithm for feature selection. It is motivated by the fact that sets with many features have a reduced accuracy during classification and recognition. Recently, researchers have used Waikato Environment for Knowledge Analysis software to reduce the dimension of the feature set. This method however is static and does not allow the users to change the feature selectors. By using the genetic algorithm for feature selection, the user will be able to change the configuration of the feature selection to improve their results. The genetic algorithm takes a set with *m* features and decreases the number of features such that the set has *n* features where *n<m*.

The researchers use the Flavia dataset and the Ionosphere dataset to show how the genetic algorithm can be used to reduce the number of features. The Flavia dataset has 100 attributes and the Ionosphere dataset has 34 attributes. The researchers then address the problem of attempting to minimize both the number of features in the feature set created using the genetic algorithm along with the classification error for both datasets. The results resulted in the use of 11% of the features or 11 features from the Flavia dataset and 17.65% or 6 features from the Ionosphere dataset. When comparing the use of the genetic algorithm to the use of the Waikato Environment for Knowledge Analysis method of selecting features, the genetic algorithm frequently resulted in a greater percent accuracy though this difference was frequently quite small. Since the genetic algorithm was just as effective as the Waikato Environment for Knowledge Analysis method and allows the user to more readily modify the commands, the genetic algorithm satisfies the initial goal of the researchers.

To truly understand why the researchers chose to employ the genetic algorithm as their method of feature selection, one must first understand what the genetic algorithm is and how it works. The Genetic Algorithm, or GA, is a population-based search method that mimics the evolution process. The process is said to mimic evolution as it takes one population and manipulates it into a new population through functions like crossovers and mutations. These functions are comparable to evolution principles such as reproduction, genetic recombination, and the principle of survival of the fittest. When comparing the evolution process to the Genetic Algorithm, we can go as far as comparing the terminology used in both processes. Genetic Algorithm terms are bit strings, features, feature value, bit position, encoded string, and decoded genotype. This can be directly compared to the evolutionary terms of chromosomes, genes, allele, locus, genotype, and phenotype. These Genetic Algorithm terms mimic their comparative evolutionary terms in the way in which they are employed by the Genetic Algorithm.

The bit strings are the fitness of the solution candidates and are evaluated using an objective or fitness function. This gives numbers which are used to rank the given bit strings taken from the dataset. The fitness function however will vary based on the problem. One fitness function which is used to optimize quadratic functions is the parabolic function . This function has constants represented by {*a, b, c*} and operates over a defined set of real and complex values. The typical Genetic Algorithm is a maximizer of the objective function. The Genetic Algorithm used by these researchers however is a minimizer. Both versions of the Genetic Algorithm however can be used as both a minimizer and maximizer as by maximizing the negative of the function, you obtain the minimum of the standard function. The Genetic Algorithm takes a population of binary bit strings and evaluates them using a fitness function. The features are then assigned ‘0’ or ‘1’ to determine whether or not it is selected. Features given an index of ‘1’ are selected and then ranked such that the top *n* are selected and survive. The elitist members of this group are automatically added to the new population. The others then undergo crossover and mutation before being added to the new population. This method results in the final features by which the data is then evaluated. Hence, by this process, the Genetic Algorithm is employed to reduce the number of features in the dataset.

Now that we understand the process employed by the Genetic Algorithm, we can look into the methods the researchers employed to be able to test their objective. The researchers used two data sets with differing numbers of features and observations as the basis for their tests. They then used both the Waikato Environment for Knowledge Analysis method and the Genetic Algorithm to find a reduced feature set for both datasets. Then using their results from both methods for reducing the number of features a dataset has, they compared the final number of features along with the accuracy of the two methods and their variations. The final results showed similar classification accuracy. However, the Genetic Algorithm features frequently outperformed the Waikato Environment for Knowledge Analysis features when entered into the same set of classifiers. Because the results yielded similar accuracies, both methods can be considered reasonable methods the reduce the number of features for a given dataset. The researchers propose that the results of the methods used in this experiment show that the Genetic Algorithm is a promising method of feature reduction primarily because of its adaptability.

Looking at the paper as a whole, we can begin to identify both its strengths and its weaknesses. The papers strengths lie in its results and the fact that they satisfy the objective of the paper as a whole. The paper gives many details on the process employed by the research along with specifying many of the algorithms used. To help make the results and processes clearer to the reader, the researchers include many tables and charts. This helps make the processes more accessible to more visual learners. The only weakness I can see in the paper is that the concepts can be difficult to follow for someone who is unfamiliar with research of this structure and methods. Also, at times, the multiple tables in a row can be difficult to follow and process. Many of these perceived weaknesses come from my lack of familiarity with research papers and topics similar to this one. Overall, the paper is a strong research paper and effectively presents its results.