CHAPTER 8

CONCLUSIONS AND SCOPE FOR FURTHER STUDY

8.1   Conclusions

In this study a new feature selection algorithm namely, Simulated Relief is presented and implemented in pest and weather data set of the cotton plant. This proposed algorithm experimentally proved its efficiency by selecting the most appropriate features which enhance the classification accuracy. Experimental study of this new feature selection algorithm is carried out in the dataset containing one hundred and fifty six instances and thirteen attributes. This agricultural data set contains two classes of pests which attack the cotton plant namely Mirid bug and Mealy bug. The focus of this research is to prove that, the classification accuracy of simulated Relief is higher than the existing Relief algorithm. Microsoft Excel and Weka tool have been aided in the computation process of this research.

This research showed the effects of using the statistical approach to machine learning to search the optimal parameter values and to determine appropriate feature subset in the Feature Selection process. It has been observed that selecting the relevant features and enhancing the classification accuracy are the two important issues that radically affect the classification tasks. Based on these issues, the solution has been proposed to select the relevant features and to enhance the classification accuracy. The operations performed in these approaches reduce the dimensionality of data sets, which in turn to allow the learning algorithms to work faster and effectively. This proposed algorithm reduces the system complexity, and it also reduces the system processing time. This, in turn, saves the computation resources. The new concept of this algorithm is the combination of different statistical measures and computation techniques with the native procedure of Relief algorithm. This algorithm

statistically guarantees to find an optimal solution, when the feature spaces are highly redundant and noisy. The Simulated Relief algorithm using Multilayer Perceptron algorithm as a classifier is found to be outperformed in the feature selection process, and it enhances the accuracy level of classification. Four different computations such as simulated Relief with Naive Bayes classifier, Simulated Relief with Multilayer Perceptron, Relief with Naive Bayes classifier, Relief with Multilayer Perceptron classifier are performed in this research, and the experimental results reveal that Simulated Relief Algorithm for Feature Selection in combination with the multilayer Perceptron classifier enhances the classification accuracy.

8.2   Scope for Further Study

In this study, Simulated Relief algorithm is carried out in agricultural data set available in the UCI machine learning repository. This can be extended to enhance the classification accuracy of medical and biological data sets. This experimental study is performed with two class problem, and the further study can be extended to multiclass problems. Further study of this research can play a critical role in many domains for classification task. This study could also be extended to continuous dataset. When this research is worked with other related techniques, it can attempt to survey some effective applications in the dynamic data mining field.