Chapter 7

Summary and Future Scope

7.1 Summary

This thesis deals with the study and applications of information theoretic measures including the measures of weighted residual information. The complete work is divided into two parts in first part we analyzed the role of entropy measures in data mining and applied them for variety of data to classify them using classification rules obtained from decision trees and in second part we studied weighted residual measure of order α and type β and have shown that this measure characterized the underlying distribution uniquely.

In Chapter 1, We discussed the different information theoretic measures, their characterization and generalizations along with weighted information measures, survival function, length based model and residual entropy. Specifically, we have concentrated upon role of entropy in data mining.

Since classification plays a big role in data mining and also it provides the base for methodology used in present work, in Chapter 2, we discussed some classification algorithms along with decision tree induced algorithms and gave an outline of research methodology used.

More output with minimum input is expected from every policy or project of development. In these types of situations, decision tree provides different mode of classification and ensures to find a right decision. By classifying a data purposely, we find some commercial valuable and potential information. In Chapter 3, we discussed the application of Renyi entropy for α=2 in ID3 algorithm to develop a decision tree, which can help in following the concept of right policy for right people.

A project is said to be successfully implemented if its target is achieved in least possible time with desired results. But, it is possible only if right decision is taken at right time. In chapter 4, the decision tree has been obtained by improved ID3 algorithm based upon generalized information theoretic measure having two parameters α =1/10 and β =1. The collected data is classified into two different classes, which helps us in developing more refined rules in comparison, if modified ID3 algorithm is used for classification. Thus, we can get more clear idea while making any final decision about implementation of, or analysis of, any implemented project.

Success of the time bound project depends upon the method of decision criteria. A method can be finalized using comparison or on the nature of training data set. In this chapter 5, the decision tree obtained by C4.5 algorithm is compared with that of improved ID3 algorithm based upon generalized information theoretic measure having two parameters α =1/10 and β =1, and found that the better results are obtained on using continuous values instead of intervals for a attribute. The collected data is classified into two different classes, which helps us in developing more refined rules in comparison, if C4.5 algorithm is used for classification. These refined rules are helpful in systematic development or progressive analysis of any system.

Weighted generalized information measures play an essential role in modeling of statistical data which includes certain amount of damaged or un-

observed values. We have proposed and studied the concept of weighted generalized residual entropy of order α and type β in Chapter 6. This proposed residual information measure characterizes the distribution function uniquely. The measure introduced here can be of interest in such type of problems.

7.2 Future Scope

As it is clear from present work entropy based information gain plays a key role during the formation of decision trees. These decision trees generated by ID3 algorithm, C4.5 algorithm etc. can induce a set of rules that helps in formation of classification criteria for any new untrained data. In future, by applying new information theoretic measures for the measurement of information gain using different values of parameters a new set of rules can be formed. Different results are obtained based upon the different decision trees for the same training data, by using different information measures. Hence, intelligent selection of a theoretic measure is also an important factor in a classification algorithm. Thus, set up of a good selection criteria/rules based on information theoretic measure can also be considered as future scope of present research work. The present work can be extended to characterize life time distribution using the concept of bivariate distribution, and further, such information theoretic measures can also be studied in order statistics.