Chapter 7 Conclusions

Many classification algorithms developed in the data mining community can only acquire knowledge on the nominal attributes' data sets. However, many real world classification tasks exist that involve continuous attributes, such that these algorithms cannot be applied unless the continuous attributes are discretized. The VPRS model is a powerful mathematical tool for data analysis and knowledge discovery from inconsistent and ambiguous data. It cannot be applied to extract rules from the continuous attributes unless they are first discretized.

In this study we first propose an extended Chi2 algorithm that determines the pre-defined misclassification rate (δ) from the data itself. We also consider the effect of variance in the two adjacent intervals. With these modifications, the extended Chi2 algorithm not only handles misclassified or uncertain data, but also becomes a completely automated discretization method and its predictive accuracy is better than the original Chi2 algorithm.

For m attributes, the computational complexity of original Chi2 algorithm at phase 1 has $O(Kmn \log n)$, where n is the number of objects in the dataset, and K is the number of incremental steps. A similar complexity can be obtained for phase 2. Although our proposed algorithm adds one step (i.e., to select the merging intervals), it does not increase the computational complexity as compared to the original Chi2 algorithm. The computational complexities of the original Chi2 algorithm, modified Chi2 algorithm, and our proposed algorithm are the same.

In addition, the VPRS model lacks a feasible method to determine a precision parameter value to control the choice of β-reducts; therefore, we propose an effective approach to select the β-reducts. First, we calculate a precision parameter value to find the subsets of an information system that is based on the least upper bound of the data misclassification error. Next, we measure the quality of classification and remove redundant attributes from each subset. Two numerical examples have been conducted to demonstrate the feasibility of the proposed approach. The implementation results show that high quality rules and a decrease in time are needed for the selection of a suitable β-reduct.

Furthermore, a real case study from communication industry shows that the VPRS theory using our proposed procedures can be applied to reduce the redundant RF functional test items in mobile phone manufacturing. VPRS also demonstrates a better performance than that of the decision tree approach. Moreover, the operation time of the RF functional test procedure by using the VPRS model is significantly less than that of the decision tree approach and the original RF functional test procedure. By using the VPRS model, the throughput will increase and the time to market will be reduced. In addition, the extracted rules constructed in this study can be used to interpret the relationship between condition and decision attributes and help companies to construct their own knowledge base for training new engineers.

Further study should develop an induction rules algorithm to induct the generation rules of VPRS.