CHAPTER IV

CONCLUSIONS

In this dissertation, we proposed a new rule extraction algorithm from a trained neural network for data mining systems not restrict to the binary value form. The rule extraction is based on the *open-box approach* and can work with numeric data. The algorithm analyzes the weight links between the input units and the hidden units of the trained neural network. In addition, the algorithm does not require any complex network structures [15, 17, 18] or need special knowledge before training like KBANN [8]. The advantages of the proposed rule extraction algorithm are that it can used with neural network software which provides the weight link values without any special network structure to extract the rule and it is not the domain specific algorithm. So that it is *portable* with the multi-layer perceptron neural network. The rule extraction process from a trained neural network can be benefit in the data mining application because it can produce new knowledge of a specific domain (application) because the neural network and the training weight links of the model can capture the relationship between the input vectors. The proposed algorithm with activation projection of the hidden node and the input vector will give the information of how the data class is classified.

As the feature extraction of data mining, the redundant neurons are possibly pruned from the REAP algorithm to get the less complex network structure and less computational time if we can eliminate some input vectors that were not use to classify

the data set. The experimental results from the rule extraction process show a comparable classification accuracy with those provided by neural networks and can be used with the real world applications [23, 24].

The rules from REAP algorithm is in the form of crisp rule projected on the input vectors. The advantages of REAP algorithm are that we receive an understandable crisp rules and can solve the problem of mathematical inequality form rules [19, 20]. The rules from RECF algorithm represent with the value of certainty factor for each rule. So that the essential information for reasoning the results is captured by the CF rule and can provide the certainty of the conclusion of the if-then rules for a problem domain. Especially in the medical application, the physician can be more certain about the given class of disease with the CF value equals to 100 than the CF value equals to 70. Finally, The NL rule that is the result of the RENL algorithm provides rules in the form of natural language terms which is readable, conceptual, and meaningful rules.

Note that the proposed rule extraction process is used for the classification problem with the supervised learning, so that the user needs to know the class of the data set. The dimension of the input data is not limited. The network for the rule extraction algorithm is equal to the number of the classes.

The proposed rule extraction process using the MLP neural network as a data mining process model efficient for the following reasons: (i) can train and retrain with no prior knowledge provided to the neural network [8], (ii) can use with a numeric data, (iii) do not depend on the problem domain, (iv) provide the certainty factor (CF) value, and (v) provide the symbolic rule as a natural language term.