# CHAPTER I

# INTRODUCTION

It is obvious that the accumulation of data by various business, scientific, and government organizations around the world is increasing daily. Thus, an intelligent data analysis processing or data mining with high speed and high accuracy for enormous volume of data is crucially required [1]. The objectives of data mining are the followings: (i) to extract implicit and useful information from data by dealing with the discovery of hidden knowledge, unexpected patterns and new rules from a large database, and (ii) to increase and improve the understandable of the data set to human using multidisciplinary fields of researches. Adriaans and Zantinqe [1] classified the data mining methods into six categories as follows: evolutionary programming, memory based reasoning, decision trees, genetic algorithm, nonlinear regression method, and neural network. Evolutionary programming automatically formulates hypothesis dependence of the target variable and the input variable. Memory-based reasoning is used to forecast a future situation while decision trees can be applied for classification tasks. Genetic algorithm is a technique for solving various combinatorial or optimization problems. Nonlinear regression is searching for the dependence of the target variable and the input variable in form of mapping function. Neural network is a knowledge discovery paradigm consisting of a massive number of neurons with a very high degree of interconnectivity.

A neural network provides a general and practical method for learning numeric data from examples. Feedforward Neural Networks (FNN) have been successfully used as a tool for classification in a variety of real-world applications [2, 3]. The development of algorithm extracts rules that is comprehensible for human from a trained feedforward neural network is crucially needed [3]. The extracted rules must be expressed in forms of if-then rule in order to make the rules readable and comprehensible.

## 1.1 Problem Reviews

Hayashi, Setiono, and Yoshida [3] applied NeuroRule to the problem of diagnosing hepatobiliary disorder while Jagielska [4] extracted knowledge for heart disease. Fu [5] and Gupta, Park, and Lam [2] reviewed a classification framework for rule extraction algorithms. In general, there are two approaches to extract rules from a trained neural network. The first approach is the *analytical approach* which extracts rules by directly interpreting the strength of the weights in the trained network. This approach is a *non-search-based* or an *open-box approach*. The examples of the open-box approach are the Cascade ARTMAP algorithm by Tan [6], M-of-N algorithm by Setiono [7], the KBANN algorithm by Shavlik [8, 9], and the Fast Extraction of Rules from Neural Network (FERNN) by Setiono and Leow [10]. The second approach is the *generate-and-test approach* which their only input/output behaviors are observed. This is a *search-based* or *black-box approach* that requires a high computational complexity of the rule extraction algorithms [2]. The example of the black-box approach is the medical diagnostic expert system by Saito and Nakano [11]. In addition, Taha and Ghosh [12] discussed some representative rule extraction techniques on Link Rule Extraction (LRE)

[7], Black-box Rule Extraction (BRE), extracting rules from fuzzy artificial neural networks [13, 14], and extracting rule from recurrent network [15].

Rule extraction can be applied to a neural network at different stages of the network construction. The extraction process, depending on when it is performed, can be classified into three types: (i) the *pre-processing type*, (ii) the *modification of neural network structure or algorithmic type*, and (iii) the *post-processing type*. For the *pre-processing type*, the prior knowledge will be provided to the neural network before the neural network training. For example, Knowledge Based Artificial Neural Network (KBANN), proposed by Towell and Shavlik [8, 9], inserted the knowledge into the network to set the bias link before training. The KBANN approach produced neural networks which it topological structure matched the dependency structure of the rules in an approximately-correct domain theory [8]. The domain theory produced a useful inductive bias by (i) focusing attention on relevant input features and (ii) indicating useful intermediate conclusion. *The modification of neural network structure or algorithmic type* concerned about the specialized network structure and training methods for extraction algorithms. This type may depend on the problem domain or may have a complex network structure such as rule extraction by successive regularization [16], fuzzy neural network [17], and Interpretable Multi-Layer Perceptron (IMLP) which used the staircase function instead of the sigmoid function [18]. There are two reasons for adopting specialized schemes, either to customize the network to a specific problem domain or to facilitate the extraction process [2]. Most often, rule extraction is applied after the network training is over. The *post-processing type* involved the existence of hidden nodes in rule extraction, so as to enhance the generalization of the algorithm [2].

This type used the output data generated from the trained neural network such as the M-of-N rule extraction algorithm [7] and the modified RX algorithm [19]. However, it is possible to deploy any combinations of these three types. For example, Setiono [7] introduced an algorithm with restriction on the inputs and weights to binary values of either minus one or plus one.

## 1.2 Statement of Problems

Current rule extraction algorithms have some limitations as follows:

1. The extracted rules are difficult to understand because they may be in forms of mathematical equations. The example of rule in form of mathematical equations is shown as follow [20].

| |
|---|
| R1: If ($A_s$–3.98$A_p$ ≥2.343) and (11.21 ≤ $A_s$–5.56$A_p$ ≤ 21.87or $A_p$–0.18$A_s$ = 1.47) |
| Then Setosa |
| where $A_s$ is sepal area and $A_p$ is petal area |

2. The extracted algorithms produce too many rules. The example of rules that produces many rules are shown as follows [19].

| | | |
|---|---|---|
| R1: | If (3.6 ≤ $a_1$ ≤ 4.2 and 85.8 ≤ $a_2$ ≤ 95.3) | Then Setosa |
| R2: | If (2.3 ≤ $a_1$ ≤ 3.5 and 69.7 ≤ $a_2$ ≤ 85.2) | Then Setosa |
| R3: | If (3.2 ≤ $a_1$ ≤ 4.5 and 82.9 ≤ $a_2$ ≤ 103.5) | Then Setosa |
| R4: | If (4.0 ≤ $a_1$ ≤ 5.4 and 95.2 ≤ $a_2$ ≤ 113.1) | Then Setosa |
| R5: | If ($a_1$ = 5.5 and $a_2$ = 111.6) | Then Setosa |
| where | $a_1 = 0.22A_s - 2.16A_p + 0.84$ | $a_2 = 3.12A_s - 11.61A_p + 41.99$ |

3. The input data format of some algorithm takes only binary values for training [12].

4. The neural network structure needs some modification for rule extraction process [16,18]. So the rule extraction algorithm may depend on the problem domain or may have complex network structure.

## 1.3 Objectives

The main objectives of this study are as follows.

1. To develop algorithm to extract rules in forms of if-then rules from a feedforward supervised neural network that project on the dimensional axis without mathematical equation in the premise of the rules.

2. To apply the rule extraction algorithm to the real world problems.

## 1.4 Scope of Work

Neural networks have been applied with success to variety of real-world problems with their learning, classification, and generalization capabilities. This dissertation is focused on solving the low degree of human comprehensibility by compiling the knowledge captured in the topology and weight matrix of a neural network into a symbolic form of if-then rules.

The rule extraction algorithm is constrained on a 3-layer MLP feedforward neural network structure. The learning process of Backpropagation is considered due to its simplicity and feasibility for rule extraction. The network employs a set of hyperplanes for classifying a given data set which is suitable to perform the interval projection process. The rule extraction process uses the open-box approach and the post- processing type.

The dissertation presents an algorithm for extracting rules from a trained supervised neural network. The input of the rule extraction process is not based on the raw data but it is based on the connection weight values given from the trained neural network [8, 9, 18, 20]. The objective of this dissertation is to find comprehensible rules to describe the behavior of the trained supervised neural network.

Each training data uses numeric input pattern and numeric target. The benchmarks for testing the rule extraction algorithm are the Glass database, the Iris database, and the Wisconsin breast cancer database. Data will be divided into test set and training set selected randomly. The input data format can be numeric data.

The performance of the rule extraction algorithm will be evaluated by three criterions [22].

1. Complexity of the algorithm. The efficiency of an algorithm is usually measured by the number of basic operations for the task (time complexity) and the amount of storage space used (space complexity). Since rule extraction methods are often based on the tests of a large number of combinations of the network inputs or parameters, the time complexity is the important factor when estimating the efficiency of the method whereas space complexity plays only the secondary role. So, the evaluation is based on the time complexity only. The time complexity of the rule extraction algorithm depend on the method used for rule extraction, correlation to the size of the neural network (i.e. the number of layers, neurons per layer, and connection), the number of training set data, the input attributes, and the values per input attribute.

2. Quality of extracted rules. The author will evaluate the quality of rules by two features [9].

   2.1 The accuracy of the extracted rules describes its abilities how the extracted rules can classify correctly under the testing examples.

   2.2 The comprehensibility of the rule system is indicated by using the following criterions: the number of extracted rules, the number of antecedents per rules [20], the overlapping of premises, and the value of the premises in forms of mathematical equations [19, 20] or in forms of input features.

3. Applicability of the algorithms. The applicability of a rule extraction algorithm considers the requirements of a method imposed on the neural networks and on the domain. The constraints of neural networks are the limitations of size and the number of layers, a special way neurons are connected (e.g. feedforward or recurrent, fully or sparsely connected), the kind of information coding in the network (local or distributed), and the special kinds of neurons (e.g. local response neurons, neurons with monotonic or approximately step activation functions). The domains to which an algorithm for rule extraction can be applied are different in sizes and the way information is provided such as the restriction on the number of domain attributes used for learning, the domain values (discrete and/or continuous), and the range of their values used to encode the learning task.

The dissertation is organized as follows. Chapter II is the theoretical background and the rule extraction process. Chapter III presents the experimental results. The discussion and conclusions are given in Chapter IV.