NANYANG TECHNOLOGICAL UNIVERSITY

A Hebbian based Rule Reduction Approach to Neuro-Fuzzy Modeling

A thesis submitted to Nanyang Technological University in fulfillment of the requirement for the degree of Doctor of Philosophy

by

Liu Feng

School of Computer Engineering

2008

Abstract

The research of neuro-fuzzy modeling is divided into two branches, the precise modeling, implemented by TSK-type fuzzy rules, and the linguistic modeling, realized through Mamdani-type fuzzy rules. Linguistic neuro-fuzzy modeling is investigated in the thesis. Neuro-fuzzy system is a hybrid of neural network and fuzzy system. Many neuro-fuzzy systems have been proposed primarily for improving modeling accuracy, but lesser attention has been devoted to address the interpretability of the derived fuzzy rules. The fuzzy rules that contain redundancy and inconsistency in the rule structure post ambiguity for people to comprehend the data and make decisions. The objective of the thesis is to construct a generic linguistic neuro-fuzzy system that is capable of generating interpretable fuzzy rules while maintaining an acceptable modeling accuracy. The major contributions of the thesis are summarized as follows.

A Hebbian based Rule Reduction (*HeRR*) neuro-fuzzy system is proposed to generate interpretable fuzzy rules. The interpretability of the rule set is significantly improved through the merger of redundant fuzzy sets and the removal of inconsistent rules. Consequently, the fuzzy membership functions can be automatically demarcated with clear semantics.

Furthermore, an iterative learning scheme is proposed to strike the balance between the interpretability of the rules and the accuracy of the modeling process; this includes

a tuning and a reduction phase. During the tuning phase, the fuzzy membership functions are adjusted using the Least Mean Squared (LMS) algorithm, while the fuzzy sets and rules are respectively merged and reduced through the HeRR in the reduction phase.

To generalize the neuro-fuzzy system to handle the pattern classification problems, the HeRR is reformulated as a min-max neuro-fuzzy classifier. A novel attribute reduction algorithm is proposed to eliminate redundancy in the rule set using rough set theory for the knowledge reduction. The algorithm is integrated into the HeRR, termed *RS-HeRR* for pattern classification. The removal of inefficient input attributes not only improves the interpretability of the rules, but may also enhance the classification accuracy, due to the clarity of the reduced rule set.

An extensive set of experiments has been undertaken to demonstrate the efficacy of HeRR and RS-HeRR. Most importantly, they have been applied to the problems of ICU artificial ventilation modeling and bank failure prediction. In contrast to other well-established neuro-fuzzy systems and classifiers, the proposed system is able to deliver superior modeling accuracy through interpretable fuzzy rules.

Acknowledgements

I will express my most gratitude and indebtedness to my supervisor **Dr. Ng Geok See**, who gives me endless supervision during the period of my Ph.D. study. I greatly benefit from his careful and rigorous attitude in research. Without his unselfish helps, I cannot finish my research work, as well as the thesis.

I gratefully acknowledge **Dr. Quek Hiok Chai**, who gave me plenty of advices during the research period, especially when I met difficulties. Without his insight in the research topic, the research cannot proceed rapidly.

Special thanks to the colleagues in the Center of Computational Intelligence (C^2I) for creating such a pleasant research environment. I would like to acknowledge **Tung Whye Loon** and **Ang Kai Keng**, whom I have discussed the ideas and directions of the research topic. Particular thanks to **Quah Kian Hong, Ting Chan Wai, Chen Jinmiao, Ma Yang, Zhu Lin, Wang Zhen, Ma Li, and Huang Haoming**. My gratitude is also given to the lab technicians: **Mr. Tan Swee Huat** and **Mr. Lau Boon Chee** for their excellent technical support.

I owe my deepest appreciation to my parents who give me the greatest emotional support and encouragement. Without them, I would not have persisted in to the end.

Finally, I would like to thank the Nanyang Technological University for offering me this precious opportunity to pursue the degree of Doctor of Philosophy.

Table of Contents

List of Figures

List of Tables

CHAPTER

1

1 Introduction

"To accomplish great things, we must not only act, but also dream; not only plan but also believe."
- Anatole France

1.1 Motivations

Soft computing (Zadeh 1994) differs from the conventional hard computing in that, it is tolerant to imprecision, uncertainty and approximation. The principal constituents of soft computing includes fuzzy logic, neural computing, evolutionary computation, machine learning and probabilistic reasoning, with the latter subsuming belief networks, chaos theory and parts of learning theory (Frank, Mario et al. 2005). Some of the relevant technical and research issues, pertaining to the development of an interpretable soft computing system, are discussed in the following sections.

1.1.1 Issues on neuro-fuzzy system

Neural network (Gerard 2005) is a learning machine designed to model brain functions and imitate human's learning capabilities. It is a nonlinear model that is able to adapt its synaptic weights to the surrounding environment. However, due to its black-box nature (Bernd and Personnaz 1999), the weights of a neural network do not possess semantic meanings and the derived knowledge after learning can neither be extracted nor comprehensible to human beings.

Fuzzy system, roots from the Zadeh's fuzzy set theory (Zadeh 1973) exploits a linguistic model to model a problem, instead of complex mathematical models (James 2005). Its strength lies in the ability to incorporate human expert knowledge and perform reasoning through a set of fuzzy rules that are readily comprehensible by the users. However, as the system becomes more complex, it is difficult to define and tune the rules manually. The rules derived from the interview with experts are usually inaccurate and biased towards the expert and not necessarily fully reflecting the relationship in data. Hence, it is desirable to formulate an automatic process to tune the rule base on the basis of data.

The hybridization of neural network and fuzzy system, called neuro-fuzzy system, combines the advantages of learning ability in the neural networks and transparency in fuzzy systems. Transparency is a measure of the human linguistic interpretability of the rules (Paiva and Dourado 2004). It allows the transformation of data information into human knowledge, which is an important property in engineering applications. The neuro-fuzzy system provides a systematic way to generate interpretable fuzzy rules and automatically tune the parameters using the neural network techniques.

1.1.2 Issues on the interpretability

Much of prevailing research on neuro-fuzzy systems has paid considerable attention on the improvement of modeling accuracy, but with lesser consideration on the interpretability issue. Interpretability refers to the ability of the fuzzy model to concisely express the behavior of the system in a comprehensible way, and accuracy refers to the capability of the fuzzy model to faithfully represent the modeled system (Casillas, Cordon et al. 2003). The generation of less-interpretable rules will greatly diminish the transparency of the fuzzy system, and make it more difficult for people to extract knowledge to analyze the data and make decisions. However, interpretability and accuracy are usually two contradictory requirements. An increase in accuracy may require changes in the fuzzy sets; consequently altering their semantics. On the other hand, strict adherence to interpretability would impose more constraints on learning and may affect accuracy. Hence, it is highly desirable to strike a balance between interpretability and accuracy.

1.1.3 Issues on the use of linguistic model

The research of neuro-fuzzy modeling can be broadly grouped as: linguistic neuro-fuzzy modeling and precise neuro-fuzzy modeling. The former focuses on models with good interpretability and performs the fuzzy inference using the Mamdani-type fuzzy rules (Mamdani and Assilian 1975). The latter concentrates on models with high accuracy and models the problem through the Takagi-Sugeno-Kang (TSK) type rules (Takagi and Sugeno 1985) (Sugeno and Kang 1988). The linguistic neuro-fuzzy modeling approach allows the problem verbally interpretable through the use of concepts such as linguistic variables (Zadeh 1975; Zadeh 1996; Herrera and

Martinez 2000; Wang and Hao 2006).

1.1.4 Issues on the drawbacks of existing neuro-fuzzy systems

The problem tackled with in the thesis is that: given a set of samples, to construct a transparent and accurate model of the data through fuzzy rules. The main objective of the thesis is to develop a systematic approach to improve interpretability of the linguistic neuro-fuzzy systems, while maintaining acceptable modeling accuracy. Currently, the major problems encountered by the existing neuro-fuzzy systems are summarized as follows:

1) *Apriority Knowledge* (Casillas, Cordon et al. 2003). Many existing neuro-fuzzy systems require the user to provide the number of the fuzzy membership functions (MFs) in each dimension. However, this task will be harder when the dimension becomes large. The size of the rule base is directly dependent on the number of MFs.
2) *Knowledge Reduction*
    i. *Inconsistent rules* (Nauck, Klawonn et al. 1997). Inconsistent rules will create confusion on the explanations of the rules. The modeling accuracy may be also adversely affected.
    ii. *Redundant attributes* (Jensen and Shen 2004). More attributes, which may not be significant, lead to more rules to be derived. This causes interpretability to be deteriorated when the number of rules becomes large.

The modeling accuracy would also suffer due to the presence of these redundant attributes.

3) *Trade-off between interpretability and accuracy* (Casillas, Cordon et al. 2003). As mentioned previously, it would be possibly contradictory to improve interpretability and accuracy simultaneously. To derive interpretable and effective rules, compromises become inevitable.

Such weaknesses in the existing neuro-fuzzy systems are direct consequents of a lack of consideration on the interpretability issue. The research effort in the thesis is heavily motivated by these points and attempts to devise solutions to these issues.

1.1.5 Issues on the use of rough set

Knowledge can be perceived as a body of information about some parts of reality, which constitute the problem domain. In neuro-fuzzy systems, the knowledge is represented as a set of attributes and fuzzy rules. The reduction of ineffiective attributes and redundant rules is essential to the improvement of interpretability in neuro-fuzzy systems, particularly for high-dimensional problems. Furthermore, the removal of redundancy does not only speed up the fuzzy inference process, but may also enhance modeling accuracy. Recently, the rough set theory (Pawlak 1991; Pawlak and Skowron 2007) provides rigorous mathematical tools for the reduction of knowledge redundancy. It potentially reduces the dimensionality of the attribute set without losing necessary information for modeling.

1.2 Organization

This thesis contains 6 Chapters. Chapter 2 introduces the neural network, the rough set theory and the fuzzy system. The design issues of neuro-fuzzy hybrid system are discussed, as well as the state-of-the-art of its hybridization with rough set. Chapter 3 focuses on the hybrid neuro-fuzzy system. The proposed Hebbian based rule reduction (HeRR) neuro-fuzzy system allows the automatic derivation of fuzzy rules based on data. The iterative tuning and reduction process is also presented to achieve a better balance between interpretability and accuracy. Chapter 4 describes the rough-set based attribute reduction algorithm and the RS-HeRR neuro-fuzzy classifier. It employs rough set approaches to reduce the knowledge without compromising on accuracy. Chapter 5 presents the results of two applications; namely the ICU artificial ventilation modeling and the bank failure prediction. Chapter 6 concludes the whole thesis.

CHAPTER

2

2 Literature Review

"If I have seen further than others,
it is by standing upon the shoulders of giants."
- Isaac Newton

The chapter presents the literature reviews. Section 2.1 provides the review on the neural networks, which is used for the learning of fuzzy membership functions and fuzzy rules in the thesis. Section 2.2 describes the fuzzy system, as well as its interpretability issue. Section 2.3 presents the neuro-fuzzy system, which is a hybrid of neural network and fuzzy system. The design issues of a neuro-fuzzy system are discussed, including the determination of fuzzy membership functions, the reduction of knowledge, and the trade-off between interpretability and accuracy. Section 2.4 introduces the fundamentals of the rough set and the existing works on the hybrid of neuro-fuzzy system with rough set, to achieve knowledge reduction.

2.1 Neural Networks

A review of the development of the neural network, as well as its advantages and disadvantages, is presented in this section.

2.1.1 Introduction

Research on neural networks is inspired by interest in the neuro-physiological fundamentals of the human brain. A neural network is a massively parallel-distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use at a subsequent time. It resembles the brain in two respects: (1) Knowledge is acquired by the network from its environment through a learning process; (2) Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

The research on the neural network is started by the pioneering work of McCulloch and Pitts (McCulloch and Pitts 1943). They described a logical calculus of the neural network model by integrating the studies of neurophysiology and mathematical logic. The next major development of the neural network is published by Hebb. He introduces the famous learning principle, which states that the effectiveness of a variable synapse between two neurons is increased by the repeated activation of one neuron by the other across that synapse (Hebb 1949). In 1954, Minsky wrote his doctoral thesis on the neural network at Princeton University (Minsky 1954). Later on, Rosenblatt introduced a new approach to the pattern recognition problem in his work on the *perceptron* (Rosenblatt 1958). In 1960, Widrow and Hoff introduced the *least mean square* (LMS) algorithm and used it to formulate the Adaline (adaptive linear element) (Widrow and Hoff 1960). An important problem encountered in the design of a multilayer perceptron is the *credit assignment problem*, termed by Minsky

(Minsky 1961). However, in 1969, a book by Minsky and Papert employed mathematics to demonstrate that there are fundamental limits on what single-layer perceptrons can compute and stated that these limits cannot be overcome by multilayer perceptrons (Minsky and Papert 1969). Later, the research on neural networks experienced its cold winter for about 10 years. It is because, first, as there were no personal computers at that time, it is difficult to conduct experiments; second, the 1969 monograph by Minsky and Papert certainly did not encourage anyone to work on perceptrons, or agencies to support the work on them. In the 1980s, there was a resurgence of interest in neural networks, due to several contributions on this field. Grossberg established a new principle of self-organization known as *adaptive resonance theory* (ART) (Grossberg 1980). In 1982, Hopfield used an energy function to formulate a new way to understand the computation performed by recurrent networks. He created a particular class of neural networks with feedbacks, which is known as *Hopfield networks* (Hopfield 1982). In 1986, the development of the *back-propagation algorithm* was proposed by Rumelhart, Hinton, and Williams (Rumelhart, Hinton et al. 1986). The two-volume book, entitled with *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, by Rumelhart and McClelland, makes it the most popular learning algorithm for training a multilayer perceptrons (Rumelhart and McClelland 1986). Later, another feedforward network with radial basis functions (RBF) is described by Broomhead and Lowe (Broomhead and Lowe 1988). It provides an alternative to multilayer perceptrons.

In recent years, the neural network has been developing at a very fast pace and have been applied to many areas. However, some difficulties still exist to model a brain-like neural network:

1. **It is difficult to find out which neurons should be connected to which.** This is the problem of determining the neural network structure. Further, the interconnections in the brain are constantly changing. The initial interconnections seem to be largely governed by genetic factors.
2. **The functioning of individual neurons may not be so simple after all.** A neuron may receive signals from many neighboring neurons and the combined stimulus may exceed its threshold. Actually the neuron need not receive all signals at exactly the same time, but must receive them all in a short time interval.

A comprehensive introduction of the neural network can be found in the books (Haykin 1999; Gerard 2005). Some of the problems mentioned above are addressed.

In the thesis, neural network techniques are used to improve the modeling accuracy. The least mean square (LMS) algorithm has been used to tune the parameters of the fuzzy membership functions. Furthermore, the Hebbian learning principle has been used in the automatic formulation of fuzzy rules (see Chapter 3).

2.1.2 Advantages and disadvantages

The strengths of the neural networks include (Haykin 1999):

1) Nonlinearity: Neural network is able to model nonlinear input/output mappings, utilizing an interconnection of nonlinear neurons. This is an important property if the underlying physical mechanism responsible for generation of the input signals is inherently nonlinear.
2) Adaptivity: Neural network can adapt their weights to the changes in the surrounding environment. It is useful in adaptive pattern classification, adaptive signal processing and adaptive control.
3) Fault tolerance: Due to the distributed nature of information storage of the neural network, the damage of a neuron or its connecting links does not degrade the performance of the whole network seriously.

The weaknesses of the neural network include:

1) Difficult to interpret (Mitra and Hayashi 2000). Neural network learns and performs through the computation by a set of neurons and connecting links. However, the weights of the links of a trained network are almost impossible to interpret and the common knowledge is difficult to extract from the network.
2) Cannot incorporate with prior knowledge (Tung 2003). As the semantic meaning of the weights is unknown, prior knowledge of the problem domain cannot be gainfully utilized by the neural network.

2.2 Fuzzy System

An introduction of the fuzzy system, as well as its strength and weakness are described in this section.

2.2.1 Introduction

A fuzzy system consists of four main components, a fuzzifier, a rule base, an inference engine and a defuzzifier, as shown in Figure 2-1. The crisp inputs X are first fuzzified as fuzzy inputs through the fuzzifier. Fuzzy inference is performed in the inference engine at the basis of a set of fuzzy rules in the rule base. It transforms the fuzzy inputs to the fuzzy outputs. These fuzzy outputs are finally defuzzified to the crisp outputs Y through the defuzzifier.

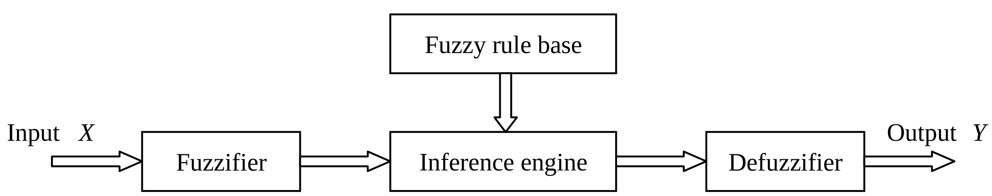


Figure 2-1: Architecture of fuzzy system.

The fuzzy system is mainly classified into three models: the Mamdani model (Mamdani and Assilian 1975), the Takagi-Sugeno-Kang (TSK) model (Takagi and Sugeno 1985) (Sugeno and Kang 1988), and the fuzzy relation equations (FRE) model (Stamou and Tzafestas 1999). The Mamdani model mainly focuses on the interpretability issue, while the TSK model aims to device a fuzzy model with a good accuracy. The modeling using Mamdani model and TSK model are usually referred as linguistic fuzzy modeling (LFM) and precise fuzzy modeling (PFM) (Casillas, Cordon et al. 2003). The fuzzy relation equations model tries to map from the input to the output space by solving FREs. This method employs FREs instead of fuzzy rules. Thus it is beyond the topic of the thesis.

In the Mamdani model, both the antecedent and the consequent parts of the fuzzy rule are fuzzy sets. An example is shown in Eq. (2-1).

$$\text{IF } x_1 \text{ is } A_1 \text{ AND } x_2 \text{ is } A_2 \text{ THEN } y \text{ is } B \tag{2-1}$$

In the TSK model, the antecedent part is the same as that of Mamdani model, but the consequent part is a linear combination of the input variables. Eq. (2-2) gives an example of it.

$$\text{IF } x_1 \text{ is } A_1 \text{ AND } x_2 \text{ is } A_2 \text{ THEN } y = b_0 + b_1 x_1 + b_2 x_2 \tag{2-2}$$

The Mamdani-type (linguistic) fuzzy system is the main focus in the thesis, as its rule expression is easier to interpret than that of the TSK-type fuzzy system. The main strength of the linguistic fuzzy models is transparency in the modeling results. Transparency is a measure of the human linguistic interpretability of the rules issued from modeling (Paiva and Dourado 2004). It allows the transformation of data into knowledge. The interpretability issues of the fuzzy systems have received increasing attention in recent years (Casillas, Cordon et al. 2003). This topic will be elaborated in Chapter 3.

In Mamdani fuzzy systems, the knowledge base contains two different information levels, the fuzzy rule semantics (in the form of fuzzy membership functions) and the linguistic rules representing the expert knowledge (Cordon, Herrera et al. 2001). As the thesis primarily concentrates on the generation of fuzzy rules, we refer the neuro-fuzzy system as a rule based system rather than a knowledge based system.

2.2.2 Advantages and disadvantages

The strengths of the fuzzy system are:

1) Interpretable to human user. The system is characterized by a set of fuzzy rules, which can be easily understood by the human user.
2) High-level reasoning. The fuzzy system is able to model the dynamics of problem using high-level IF-THEN fuzzy rules. It can approximate the actions of an expert by its inference that simulates the human cognitive process.
3) It is easy to incorporate prior knowledge. Any domain knowledge in the form of fuzzy rules can be added into the system, without affecting the existing rules.

The drawbacks of the fuzzy system are:

1) It is not easy to manually formulate the fuzzy rule base. Fuzzy rules are derived through interviews with experts in traditional fuzzy system. However, they may be inaccurate and differ with different experts.
2) It is not easy to manually tune the parameters. For a complex system with many variables, it is difficult to tune the parameters of the system manually to boost the performance of the system.

2.3 Neuro-fuzzy hybrid system

2.3.1 Introduction

As mentioned above, the neural network is superior in its learning ability but is difficult to interpret due to its black-box nature. Fuzzy system makes use of interpretable IF-THEN fuzzy rules to model a problem domain but cannot self-adjust its parameters. The hybrid of neural network and fuzzy system, called neuro-fuzzy system, combines the advantages of fuzzy system in terms of interpretability with the advantages of neural networks regarding learning ability (Yamakawa, Uchino et al. 1992; Furuhashi, Hasegawa et al. 1993; Kasabov 1996). It usually has a five-layer network structure, shown in Figure 2-2. The layers are characterized by the fuzzy operations, which fulfill the functions of the four components of fuzzy system. They are namely: the input layer (layer one), the condition layer (layer two), the rule node layer (layer three), the consequence layer (layer four), and the output layer (layer five).

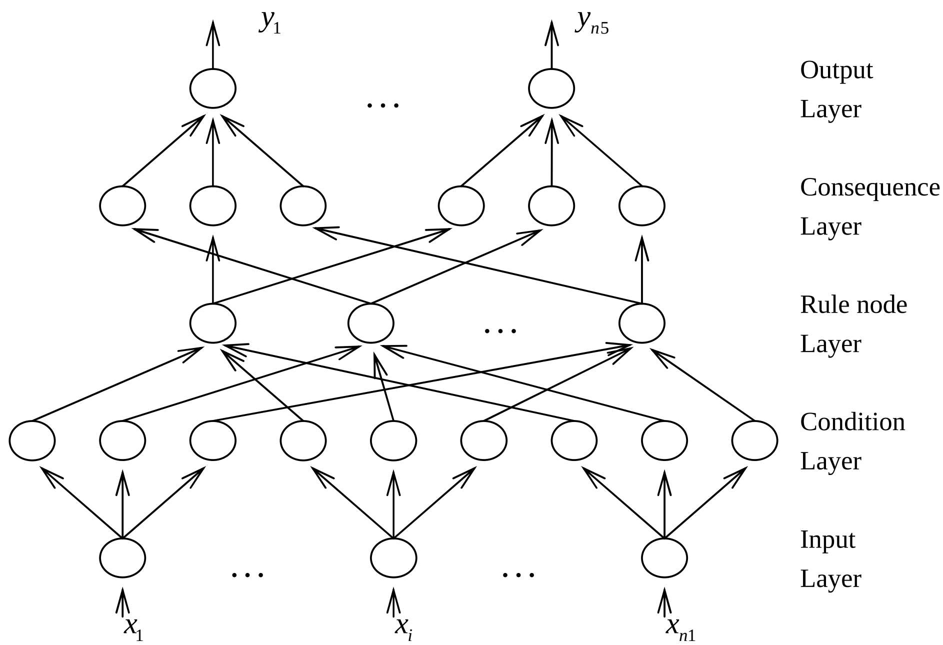


Figure 2-2: The structure of 5-layer neuro-fuzzy system

1) The input layer: The nodes of the input layer directly transfers the input signals into the second layer. The number of the input nodes is equal to the number of input variables.

2) The condition layer: The nodes of the condition layer conduct the fuzzification operation for each input variable. Each node of the condition layer corresponds to a fuzzy label for an input variable.

3) The rule node layer: Rule node layer is the core part of the neuro-fuzzy system. Each rule node links the nodes in the condition and consequence layer to formulate a fuzzy rule.

4) The consequence layer: Each node in the consequence corresponds to a fuzzy label of an output variable. It computes the membership value of each target

output.

5) The output layer: The crisp output is produced in this layer through the defuzzification operation.

In relation to neural biological, a neuro-fuzzy system has many functional similarities with the hippocampus and the basal ganglia.

The hippocampus is a part of the forebrain, located in the medial temporal lobe. It forms a part of the limbic system and plays a part in long term memory and spatial navigation. Two critical functional properties of the hippocampus system are *pattern separation* and *pattern completion*. Pattern separation leads to the formation of different, relatively non-overlapping representations in the hippocampus, while the pattern completion enables partial cues to trigger the activation of a complete, previously encoded memory. The main mechanism that the hippocampus uses to achieve pattern separation is to make the representations sparser (having fewer units active) (O'Reilly and Munakata 2000). In neuro-fuzzy system, knowledge is encoded and represented in term of rule nodes (the rule node layer). In contrast to conventional neural network like MLP, where adjacent layers are fully connected, the rule nodes in neuro-fuzzy system only link to a small set of nodes in condition and consequence layers. Thus it is less likely for the rule nodes to interfere with each other, which leads to the pattern separation. When new patterns come, the retrieval of existing memory is fulfilled through fuzzy inference, which leads to the pattern completion.

The basal ganglia are a group of nuclei in the brain interconnected with the cerebral cortex, thalamus and brainstem. Basal ganglia are associated with a variety of functions such as motor control, cognition, emotions and learning. The cortical areas project to the *striatum*, part of the basal ganglia, which performs a pattern matching function. It projects to a number of small regions know collectively as the *pallidum*. The projections to the pallidum are substantially inhibitory, and these regions in turn inhibit cells in the *thalamus*, which projects to select actions in the cortex. In other words, the pallidum serves a selection function, and the thalamus controls the execution of production actions (Anderson, Bothell et al. 2004). In neuro-fuzzy system, pattern matching is performed in the condition layer. After pattern matching, one or several rule nodes are selected in the rule node layer based on the firing strength. The output is computed through the consequence layer, which is similar to the role of thalamus.

2.3.2 The design issues of neuro-fuzzy system

Three issues in the design of a neuro-fuzzy system are discussed in this section, namely, (1) formulation of fuzzy membership functions (MFs); (2) knowledge reduction; and (3) balance between interpretability and accuracy. Section 2.3.2.1 presents the methods to partition the input/output space and form the fuzzy membership functions. Section 2.3.2.2 considers the reduction of knowledge, including the removal of inconsistent rules and redundant attributes. Section 2.3.2.3 focuses on the trade-ff between the interpretability of the derived fuzzy rules and the improvement of the modeling accuracy.

2.3.2.1 Formulation of fuzzy membership functions

For a neuro-fuzzy system, the fuzzy membership functions are formulated through the partitioning of input and output spaces. It includes the determination of the number and positions of the MFs in each dimension. The number and locations of the MFs affect both the system's modeling accuracy and the interpretability of the derived fuzzy rules. That is, the larger the number of MFs, the more the rules will be derived and the more complex the model will be obtained. As human cannot process efficiently more than 7 ± 2 different entities in the short term memory (Miller, Galanter et al. 1960), the number of MFs of each dimension should not exceed this number. On the other hand, the positions of the MFs should be placed properly that all the MFs are clearly separated with less overlap.

Many existing neuro-fuzzy models incorporate prior knowledge provided by the user to formulate the MFs, i.e. the number and the distribution of the MFs. However, as the dimension of the problem increases, it is difficult to set them manually. An automatic way of formulating the MFs is desirable.

The most direct way is to partition the input and output space into grid types. Equally separated partitions are used in (Cordon and Herrera 2000) (Kasabov 2001). Equal partitioning is easy to use due to its simplicity. The prior assumption of this partition method is the equal distribution of the data in each dimension. However, this assumption is usually unsatisfied.

The second method is to use the Genetic Algorithm (GA) to determine the proper partitions (Lee and Teng 2001) (Kaya and Alhaji 2006). The main disadvantage of this method is that it is very time consuming. Moreover, the semantics of the resultant fuzzy sets may be altered.

The third method, which is commonly used to generate the MFs, is the clustering algorithms (Baraldi and Blonda 1999; Baraldi and Blonda 1999; Xu and Wunsch II 2005). There are two ways to employ these clustering algorithms. One is that, the clustering algorithm is initially performed in the multi-dimension space, and the derived clusters are projected onto each single dimension. A disadvantage of this method is that, the projected clusters in single dimension may have a great deal of overlap, although they are well separated in a multi-dimensional space. The authors in (Juang and Lin 1998) proposed an aligned clustering based method to deal with this problem. Another approach is to perform clustering algorithms in each individual dimension (Tung and Quek 2002; Ang, Quek et al. 2003). An advantage of this method is that, the derived MFs can be clearly separated in each input dimension and easily interpreted to the user. However, the resultant partitioning usually serves as an initial partitioning and may be altered in the rule tuning process. The clustering methods often require the user to provide the number of clusters. It cannot be easily estimated.

Another method is to derive the space partitions through the merger of the fuzzy sets. Similarity measures are used to simplify the rule base (Setnes, Babuska et al. 1998).

The space partitioning is obtained simultaneously. An advantage is that, the fuzzy sets in each dimension are with little overlap, and these partitions will not be altered again. The disadvantage is that the merger of MFs may cause the decrease of the modeling accuracy. Thus, the maintaining of accuracy is needed.

2.3.2.2 Knowledge reduction

A fundamental problem of a knowledge-based system is whether the whole knowledge is always necessary to define the behavior of the system. For a neuro-fuzzy system, where the knowledge is represented as the IF-THEN fuzzy rules, the reduction of knowledge is important for the improvement of the interpretability of the rule set. It consists of two aspects, the reduction of inconsistent rules and the removal of redundant attributes.

Inconsistent rule reduction

Various methods have been proposed for the reduction or resolution of the inconsistency in the fuzzy rule-based system. Wu proposed a verification detection method for the inconsistent rules by token passing (Wu 1997). Lin proposed a reasonable logic inference approach to keep the fuzzy reasoning consistent (Lin 1996). In addition, Lin also proposed a method to merge multiple knowledge bases with conflicting rules through a process called weighted determination (Lin 1996). The necessity support and possibility support are applied to detect and remove inconsistencies (Lee, Chen et al. 2002).

Redundant attribution reduction

For a high-dimensional problem with a large number of input variables, the fuzzy system suffers from an exponential growth in its size due to the homogeneous partitioning of the input and output spaces caused by the use of linguistic variables (Bastian 1994). Even for problem whose dimensions are not large, the reduction of attributes can simplify the rule set, make the resultant model more interpretable and even boost its performance.

1) Attribute reduction outside the model

   Prior to the learning process of the neuro-fuzzy system, the attributes can be initially selected through feature selection methods (Molina, Belanche et al. 2002). Feature selection methods are divided into two categories: (1) the filter scheme, which takes place before the use of the learning algorithm (i.e. the classifier), and serves as a preprocessing step. The RELIEF algorithm (Kira and Rendell 1992) and FOCUS algorithm (Almuallim and Dietterich 1994) are two well-known filter-based feature selection methods; (2) the wrapper scheme, which uses the learning algorithm as a subroutine to evaluate the selected feature set (John, Kohavi et al. 1994; Kohavi and John 1997).

2) Attribute reduction within the model

   In this case, the attribute selection and reduction process is embedded in the neuro-fuzzy model. The task involves selecting a subset of input variables to be used in the model or removing those input variables that do not significantly contribute to its performance. Various methods has been proposed (Hong and Lee 1999) (Silipo and Berthold 2000) (Jin 2000) (Lee, Chen et al. 2001) (Casillas, Cordon et al. 2001) (Shen and Jensen 2004) (Chakraborty and Pal 2004).

3) Attribute reduction in each rule

Another approach is to select attribute subset for each derived fuzzy rule. A specific input variable ignored in a rule may be used in another one. The most significant attributes are obtained for each rule during the learning process (Castro, Castro-Schez et al. 1999) (Gonzalez and Perez 2001). Ang and Quek address the knowledge reduction through attribute reduction for the whole rule set as well as each fuzzy rule (Ang and Quek 2005).

2.3.2.3 Balance between interpretability and accuracy

The research on the trade-off between interpretability and accuracy is divided along two threads. The first thread focuses on the improvement of interpretability while maintaining acceptable accuracy (Casillas, Cordon et al. 2003), and the second thread pays more attention to the enhancement of accuracy while preserving the interpretability through certain constraints (Casillas, Cordon et al. 2003).

In the first thread, the interpretability can be improved by the removal of redundant attributes and rules. There are two approaches to the reduction of redundant rules. The first approach is to select important rules, either by GA (Ishibuchi, Nozaki et al. 1995;

Ishibuchi, Murata et al. 1997) or orthogonal transform based methods (Yen and Wang 1999; Setnes and Babuska 2001). The second approach is to merge similar or neighboring rules (Klose, Nurnberger et al. 1998; Setnes, Babuska et al. 1998; Setnes, Babuska et al. 1998). Another way to improve the interpretability is to use other rule expressions, such as the disjunctive norm for rules (Gonzalez and Perez 1998), exception rules (Krone and Kiendl 1994) and union-rule configuration (Combs and Andrews 1998).

In the second thread, one way to improve the accuracy is to employ more sophisticated rule-base learning methods. The COR (Casillas, Cordon et al. 2002) is such a method that tries to make a better cooperation among the linguistic rules. Another way is to extend the usual linguistic model structure to make it more flexible. The linguistic modifiers (Zadeh 1972; Bouchon-Meunier and Jia 1992), the double-consequent rules (Cordon and Herrera 2000; Espinosa and Vandewalle 2000), the weighted rules (Pal and Pal 1999; Ishibuchi and Nakashima 2001; Setnes, Casillas et al. 2002) and the hierarchical knowledge bases (Ishibuchi, Nozaki et al. 1995; Cordon, Herrera et al. 2002) have been used to improve the accuracy.

2.3.3 Existing neuro-fuzzy systems

The concept of the neuro-fuzzy system is rooted in the pioneering work of H. Takagi and I. Hayashi, who obtained a basic patent in Japan in 1988. The basic idea underlying their patent is to exploit the learning capability of neural network to improve the performance of fuzzy systems. In the early 1990s, two well-known neuro-fuzzy systems, the ANFIS (Jang 1993) and the FALCON (Lin and Lee 1991), were proposed. Their ideas still have much impact on the development of the neuro-fuzzy systems now. In the last decade, the neuro-fuzzy system developed rapidly. A comprehensive review of them is in (Mitra and Hayashi 2000). In this section, some neuro-fuzzy models proposed in recent years are reviewed.

The Evolving Connectionist System (ECOS) is proposed by Kasabov (Kasabov 2003). Some neuro-fuzzy systems have been proposed under this framework. The Evolving Fuzzy Neural Network (EFuNN) (Kasabov 2001) performs the fuzzy reasoning using Mamdani-type fuzzy rules, while the Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS) (Kasabov and Song 2002) and the Hybrid Neuro-Fuzzy Inference Systems (HyFIS) (Kim and Kasabov 1999) are TSK-type fuzzy systems. The ideas of DENFIS is also extended to the transductive reasoning (Song and Kasabov 2005).

Another neuro-fuzzy model is the Self-Constructing Neural Fuzzy Inference Network (SONFIN) (Juang and Lin 1998). Subsequently, it is extended with recurrent feedbacks, namely the Recurrent Self-Organizing Neural Fuzzy Inference Network (RSONFIN) (Juang and Lin 1999). The rules in RSONFIN are Mamdani-type fuzzy rules. To achieve a better accuracy, the model is also extended to the TSK-type rules (Juang 2002).

The pseudo outer-product based fuzzy neural network (POPFNN) is proposed by Zhou and Quek (Zhou and Quek 1996), based on the Hebbian's learning algorithm (Hebb 1949). It has been extended using different fuzzy inference schemes, including

POPFNN-AARS(S) (Quek and Zhou 1999), POPFNN-CRI(S) (Ang, Quek et al. 2003), POP-Yager (Quek and Singh 2005), and POPFNN-AARS(NS) (Quek and Zhou 2006). Subsequently, the rough set-based pseudo outer-product fuzzy rule identification algorithm (RSPOP) was proposed by Ang and Quek (Ang and Quek 2005) to reduce the rules derived by the POP learning algorithm using rough set theory. The generic self-organizing fuzzy neural network (GenSoFNN) (Tung and Quek 2002) is another neuro-fuzzy system based on the POPFNN family.

A dynamic fuzzy neural network (D-FNN) is proposed (Wu and Er 2000), which is a TSK-type fuzzy system based on extended radial basis function (RBF) neural networks. A hierarchical on-line self-organizing learning algorithm is proposed to derive fuzzy rules. Subsequently, the D-FNN is generalized through a fast approach for automatically generating fuzzy rules (Wu and Er 2001). Another learning algorithm for the D-FNN is proposed, in which the structure of the system can be self-adaptively determined without partitioning the input space a priori (Er and Wu 2002).

The flexible neuro-fuzzy inference system (FLEXNFIS) is proposed not only to tune the parameters of the membership functions but also the type of the systems, either Mamdani or logical (Rutkowski and Cpalka 2003). In their later work, the use of adjustable quasi-triangular norms makes the neuro-fuzzy system more flexible (Rutkowski and Cpalka 2005).

A comparison of the above-mentioned neuro-fuzzy models is shown in Table 2-1.

From the table, it can be seen that, more than half of the systems require the prior knowledge to determine the membership functions, and nearly half of the systems does not considered the reduction of rules. None of them except the RSPOP has paid attention to the attribute reduction and only the RSPOP and GenSoFNN have considered the balance of interpretability and accuracy. This comparison indicates that most of these neuro-fuzzy systems lack the consideration of interpretability issues. It motivates the research work of the thesis.

Models	**Type**	**PK**	**RR**	**AR**	**BAI**
ANFIS	TSK	×	×	×	×
FALCON	Mamdani	×	×	×	×
EFuNN	Mamdani	×	√	×	×
DENFIS	TSK	√	×	×	×
HyFIS	TSK	×	√	×	×
POPFNN	Mamdani	×	√	×	×
RSPOP	Mamdani	√	√	√	√
GenSoFNN	Mamdani	√	√	×	√
SONFIN	TSK	√	×	×	×
D-FNN	TSK	×	√	×	×
FLEXNFIS	Mamdani	×	×	×	×

Table 2-1: Comparison of neuro-fuzzy models in the literature (PK: whether it does not require prior knowledge to define the MFs, i.e. the number of MFs; RR: whether it does the rule reduction; AR: whether it does the attribute reduction inside the model; BAI: whether it considers the balance of accuracy and interpretability).

2.4 Rough Set

Rough set theory was introduced by Pawlak to deal with incomplete or vague knowledge (Pawlak 1991). A review of rough set theory can be found in (Pawlak and Skowron 2007). The strength of rough set is its ability to reduce the knowledge redundancy in a problem-independent manner. In the neuro-fuzzy modeling domain, rough set provides a rigorous mathematical tool to remove redundancy from the fuzzy rule set.

2.4.1 Fundamentals of rough set

A knowledge base can be represented as a relation system $K=(\mathbf{U},\mathbf{R})$, where $\mathbf{U}\neq\varnothing$, called *the universe,* is a set of objects, and $\mathbf{R}$ is a set of relations. In a rule-based knowledge base, the $\mathbf{R}$ particularly consists of a set of attributes, each of which forms an equivalence relation. For any $\mathbf{P}\subseteq\mathbf{R}$, an associated equivalence relation, called indiscernibility relation over P, denoted by $IND(\mathbf{P})$, is formed in Eq. (2-3):

$$IND(\mathbf{P})=\left\{(x,y)\in U^2 \mid \forall a\in\mathbf{P}, a(x)=a(y)\right\} \tag{2-3}$$

The family of all equivalence classes of the equivalence relation $IND(\mathbf{P})$ is denoted by $\mathbf{U}/IND(\mathbf{P})$.

✧ Lower and upper approximation

With each subset $X\subseteq U$ and an equivalence relation R, the rough set theory uses the R-lower and R-upper approximation to approximate X, as described in Eqs. (2-4) (2-5).

$$\underline{R}X=\bigcup\left\{Y\in\mathbf{U}/R \mid Y\subseteq X\right\} \tag{2-4}$$

$$\overline{R}X=\bigcup\left\{Y\in\mathbf{U}/R \mid Y\cap X\neq\varnothing\right\} \tag{2-5}$$

By employing the knowledge R, the $\underline{R}X$ is the set of all objects of $\mathbf{U}$ which can be classified with certainty as the elements of X, and the $\overline{R}X$ is the set of objects of $\mathbf{U}$ which can be possibly classified as elements of X.

Let P and Q be two equivalence relations in $\mathbf{U}$, the positive, negative and boundary regions are defined by Eqs. (2-6)-(2-8):

$$POS_P(Q) = \bigcup_{X \in \mathbf{U}/Q} \underline{P}X \tag{2-6}$$

$$NEG_P(Q) = \mathbf{U} - \bigcup_{X \in \mathbf{U}/Q} \overline{P}X \tag{2-7}$$

$$BND_P(Q) = \bigcup_{X \in \mathbf{U}/Q} \overline{P}X - \bigcup_{X \in \mathbf{U}/Q} \underline{P}X \tag{2-8}$$

✧ Reduct and core

In the reduction of knowledge, there are two fundamental concepts in rough set theory, *the reduct* and *the core*. A reduct of knowledge is its essential part that suffices to define all basic concepts occurring in the considered knowledge, whereas the core is the most important part of the knowledge. Considering a set of attributes (a family of equivalence relations) $\mathbf{R}$ and an attribute $R \in \mathbf{R}$, R is dispensable in $\mathbf{R}$ if it satisfies $IND(\mathbf{R}) = IND(\mathbf{R} - \{R\})$, otherwise R is indispensable in $\mathbf{R}$. $\mathbf{R}$ is independent if each $R \in \mathbf{R}$ is indispensable in $\mathbf{R}$. $\mathbf{Q} \subseteq \mathbf{P}$ is *a reduct* of $\mathbf{Q}$ if $\mathbf{Q}$ is independent and $IND(\mathbf{Q}) = IND(\mathbf{P})$. A core of $\mathbf{P}$ is the set of all indispensable relations in $\mathbf{P}$. That is, described in Eq. (2-9),

$$CORE(\mathbf{P}) = \bigcap RED(\mathbf{P}) \tag{2-9}$$

where $CORE(\mathbf{P})$ and $RED(\mathbf{P})$ denote the core and the reduct of $\mathbf{P}$ respectively.

✧ Dependency of attributes

For a knowledge base, dependency between attributes may exist. Let $\mathbf{P}, \mathbf{Q} \subseteq \mathbf{R}$. $\mathbf{Q}$ depends on $\mathbf{P}$ iff $IND(\mathbf{P}) \subseteq IND(\mathbf{Q})$. The dependency can also be partial which means that only part of $\mathbf{Q}$ depends on $\mathbf{P}$. For a partial dependency, $\mathbf{Q}$ depends on $\mathbf{P}$ in a degree k $(0 \leq k \leq 1)$, described in Eq. (2-10)

$$k = \gamma_{\mathbf{P}}(\mathbf{Q}) = \frac{\text{card } POS_{\mathbf{P}}(\mathbf{Q})}{\text{card } \mathbf{U}} \tag{2-10}$$

where $\text{card } POS_{\mathbf{P}}(\mathbf{Q})$ and $\text{card } \mathbf{U}$ denote the number of elements of the $POS_{\mathbf{P}}(\mathbf{Q})$ and $\mathbf{U}$ respectively.

2.4.2 Hybrid of neuro-fuzzy system and rough set

One of the neural fuzzy systems that incorporate the rough set theory is the rough-fuzzy MLP (Banerjee, Mitra et al. 1998). In the rough-fuzzy MLP, the domain knowledge is encoded in the connection weights and rules are generated from the decision table by computing the relative reducts. This model is further pursued using a modular approach (Mitra, Mitra et al. 2001; Pal, Mitra et al. 2003). A multi-class problem is divided into multiple 2-class sub-problems. The rough-fuzzy MLP modules, generated for each class, are combined together to automatically formulate fuzzy rules using the Genetic Algorithm (GA).

Another rough-fuzzy approach for the identification of fuzzy rules is proposed by

Shen and Chouchoulas (Shen and Chouchoulas 2002). It uses the rough set based attribute reduction algorithm (QuickReduct) to reduce the attributes of fuzzy rules derived from the rule induction algorithm (RIA) (Lozowski, Cholewo et al. 1996). This approach attempts to compute a minimal reduct without exhaustively generating all possible subsets of attributes. However, it does not always guarantee to generate a minimal reduct. Further improvement for this approach is proposed by Shen and Jensen (Shen and Jensen 2004). It uses fuzzy-rough set (Pal and Pal 1999) to better guide the attributes selection.

Recently, a hybrid of fuzzy neural network and the rough set, called the RSPOP, is proposed (Ang and Quek 2005). The RSPOP algorithm utilizes the rough set theory to perform the reduction of the attributes of the fuzzy rules generated by the pseudo outer-product (POP) rule identification algorithm (Zhou and Quek 1996; Quek and Zhou 1999; Ang, Quek et al. 2003).

2.5 Summary

In this chapter, the state-of-arts of neural network and fuzzy system are reviewed. The hybrid of neural network and fuzzy system, namely neuro-fuzzy system, is presented and reviewed. The following design issues of neuro-fuzzy system are discussed: (1) the determination of afuzzy membership functions; (2) the reduction of knowledge; (3) the trade-off between interpretability and accuracy. For knowledge reduction, the fundamentals of rough set and the existing works on its use in neuro-fuzzy system are introduced and reviewed.

In the following chapters, several methods will be proposed to construct a generic linguistic neuro-fuzzy system that is able to generate interpretable fuzzy rules while maintaining acceptable modeling accuracy.

CHAPTER

3

3 Hebbian Ordering based Rule Reduction Neuro-Fuzzy System

"It's not that I'm so smart,
it's just that I stay with problems longer."
- Albert Einstein

The Hebbian based Rule Reduction (HeRR) neuro-fuzzy system is proposed in this Chapter, to improve the interpretability of fuzzy rules. The Hebbian ordering is defined to indicate the importance of each rule. It is used to merge the fuzzy membership functions and remove inconsistent rules. An iterative tuning and reduction process is proposed to strike a trade-off between interpretability and accuracy. It reduces redundant rules and obtains more compact and meaningful rules, while still maintaining high accuracy. The detail of the algorithm is presented in Section 3.1. Section 3.2 provides five experiments to demonstrate the effectiveness of the proposed method. Conclusion is described in Section 3.3.

3.1 Hebbian based rule reduction algorithm

In the fuzzy system, the crisp inputs are firstly fuzzified to fuzzy inputs and subsequently transformed into fuzzy outputs through a set of fuzzy rules, which has the form (Mamdani type) as described in Eq. (3-1):

$$R^l: \quad \text{IF} \quad x_1 \text{ is } A \text{ and } x_2 \text{ is } B \quad \text{then } y_1 \text{ is } C \tag{3-1}$$

The fuzzy outputs are finally defuzzified into crisp outputs.

When integrating with the neural network to utilize its learning ability, a five-layer neural network is deployed, as shown in Figure 2-2. It consists of the input layer, condition layer (which performs the fuzzification), rule node layer (each node in which denotes a fuzzy rule), consequence layer (which performs the defuzzification). The crisp input and output vectors are represented as $\mathbf{X}^{\mathrm{T}}=[x_1, x_2, \ldots, x_i, \ldots, x_{n1}]$ and $\mathbf{Y}^{\mathrm{T}}=[y_i, y_2, \ldots y_i, \ldots y_{n5}]$ respectively. The terms n_1, n_2, n_3, n_4, n_5 denote the number of the neurons of the input, condition, rule-node, consequence and output layers, respectively. $n_2=\sum_{i=1}^{n_1} L_i$ and $n_4=\sum_{m=1}^{n_5} T_m$, where L_i and T_m are the number of input and output linguistic labels for each dimension. In the proposed method, the Gaussian membership function is used in the condition and consequence layers to perform the fuzzification and defuzzification. The centroids and widths of the MFs are denoted as $(c_{i,j}^{II}, \delta_{i,j}^{II})$ (the i-th input dimension and the j-th MF) and $(c_{l,m}^{IV}, \delta_{l,m}^{IV})$ (the m-th output dimension and the l-th MF), for the two layers. By denoting $IL_k(i)$ as the input label in the i-th input dimension of the k-th rule, and $OL_k(m)$ as the output label in the m-th output dimension of the k-the rule, the final output of the m-th output dimension, denoted as o_m^V, can be expressed as shown in Eq. (3-2).

$$o_m^V = \frac{\sum_{k=1}^{n_3} c_{OL_k(m),m}^{IV} \Big/ \delta_{OL_k(m),m}^{IV} \times f_k^{III}}{\sum_{k=1}^{n_3} f_k^{III} \Big/ \delta_{OL_k(m),m}^{IV}} \tag{3-2}$$

where $f_k^{III} = \prod_{i=1}^{n_1} \exp\left\{-\frac{\left(x_i - c_{i,IL_k(i)}^{II}\right)^2}{\left(\delta_{i,IL_k(i)}^{II}\right)^2}\right\}$ is called the firing strength of the input point $\mathbf{X}^{\mathrm{T}}$.

There are two important issues in neuro-fuzzy modeling, namely: the interpretability and accuracy. Interpretability refers to the capability of the fuzzy model to express the behavior of the system in an understandable way, while the accuracy refers to the capability of the fuzzy model to faithfully represent the system (Casillas, Cordon et al. 2003). Interpretability and accuracy are usually pursued for contradictory purposes and sometimes are dipoles apart as the system complexity increases. When tuning the membership functions of the rules to diminish the modeling error, the interpretability of the rules may be degraded during the tuning process, where the fuzzy sets can drift closer to each other and may end up overlapping each others (Setnes, Babuska et al. 1998). Thus, a balance between interpretability and accuracy is desirable. The purpose of the chapter is to identify and devise a way to obtain low modeling error while at the same time still maintaining high interpretability.

The proposed method consists of three phases: *the initial rule generation* **(**P1**)**, *the iterative rule reduction and refinement* (P2), and *the membership function tuning* (P3), as shown in Figure 3-1.

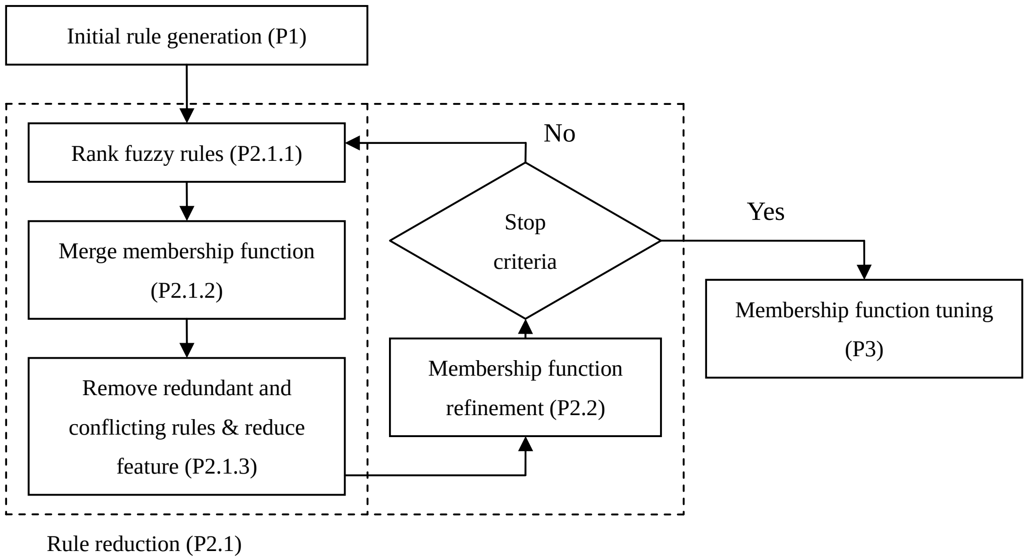


Figure 3-1: The flow chart of the proposed rule generation, reduction and tuning process.

In P1 phase, the fuzzy rules are formulated to cover all the training samples. The number of input labels, as well as the output labels in the neuro-fuzzy model, is equal to the number of the fuzzy rules. There will be an excessive set of adjustable parameters and the resultant fuzzy sets will have large areas of overlap. These will be reduced in P2 phase.

In P2 phase, an iterative tuning process is employed. It consists of two sub-phases: *rule reduction* (P2.1) and *membership function (MF) refinement* (P2.2). The objective of this phase is to improve the interpretability of the model and simultaneously to maintain high modeling accuracy. Membership functions are merged according to the degree of their overlap, and redundant and conflicting rules are deleted at the same time, in the rule reduction sub-phase. A merging scheme based on the Hebbian ordering of rules is proposed. There are 3 procedures in P2.1 phase: *rank the rules*

(P2.1.1), *merge the MFs* (P2.1.2), and *reduce the redundant and conflictive rules* (P2.1.3). The system first ranks all the rules according to the importance for each rule, which is realized by the Hebbian ordering in the proposed method. Then the MFs of each variable are merged in accordance to a set of criteria, which will be defined later. Finally, if there are any equivalent rules that have the same condition and consequence parts with others, only one of them is preserved, and if there are any conflicting rules that have the same condition but different consequence part with others, the ones with lower importance are removed. If there is only one MF of an input feature variable, this feature will be reduced. In P2.2 phase, the Least Mean Square (LMS) algorithm is employed to tune the centers and widths of the membership functions to reduce the modeling error. As there may be unsatisfied overlaps between MFs, it will iterate in P2.1 phase to reduce the MFs and the rules. The stopping criterion for iteration is specified as follows: Denote the training error after the i^{th} time of rule reduction (P2.1) and MF refinement (P2.2), as E_{train}^{i}. If i exceeds the maximum iteration $i_{\max}$, or $E_{train}^{i+1} > \eta E_{train}^{i}$, then restore the rule set just after the i^{th} iteration and goes to the next phase. η is a human defined parameter that controls the reduction of the system complexity, based on the fact that if the number of rule is less than necessary, the training error will become much larger than before.

The third phase is the finer tuning of the MFs to achieve a high level of accuracy. There will be no further reduction of the MFs and rules, and have a larger number of updating epochs and a smaller learning rate than that in the MF refinement (P2.2) of the second phase.

3.1.1 Initial rule base formulation

During the rule initialization phase, whenever a new data sample is presented, if there are no rules in the rule base, or if the strength of the rule with the largest firing strength is below a specified threshold θ, a new rule node will be created. This threshold controls the coverage of the data space by the rules and affects the number of initial rules. The larger the threshold the less number of the rules will be generated, and vice versa. When a new rule node is created, for each input and output dimension, a neuron is inserted into the condition and consequence layers. The centroid of the newly generated MFs is set to the data sample, while the width of the MFs is set to a predefined value in proportion to the scale of each dimension. The flowchart of the procedure is described in Figure 3-2.

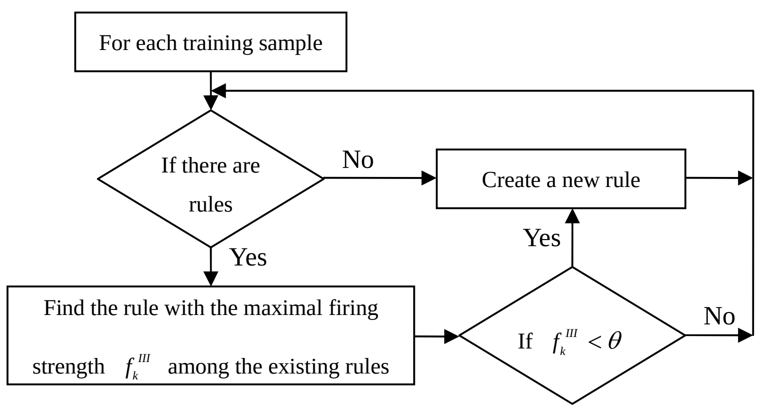


Figure 3-2: Flowchart of the rule initialization algorithm.

3.1.2 Hebbian based MF and rule reduction

Excessive MFs and rules would be generated during the rule initialization phase and they should be reduced so as to provide a clear linguistic semantic meaning of each fuzzy set and decrease the complexity of the model. In fuzzy modeling, each fuzzy rule corresponds to a sub-region of the decision space. Some rules lying in a proper region may represent many samples and have much influence on the final result, while some other rules may occasionally be generated by noise and become redundant in the rule base. Thus, the importance of each rule is not necessarily of the same value. This section introduces the *Hebbian importance* of fuzzy rules and explains how it is utilized to merge the fuzzy sets.

As the membership functions of a rule are determined, the training sample $\left(\mathbf{X}_{\mathrm{i}}^{\mathrm{T}}, \mathbf{Y}_{\mathrm{i}}^{\mathrm{T}}\right)$ is fed into both the input and output layers simultaneously. The input $\mathbf{X}_{\mathrm{i}}^{\mathrm{T}}$ is used to produce the firing strength of the k-th rule node, while $\mathbf{Y}_{\mathrm{i}}^{\mathrm{T}}$ is fed into the consequence layer to derive the membership values at the output label nodes. As stated in the Hebbian learning algorithm (Hebb 1949): "*When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes place in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.*", if the input-output samples repeatedly fire a rule by the product of their firing strength and the membership values, and the accumulated strength surpasses that of other rules, it indicates the existence of such a rule. In another view, the accumulated strength exerted by the sample pairs reflects the degree of coverage of them by the rule. The rule that covers the samples to a higher degree will have greater influence on the modeling accuracy. This is due to the fact that when the MFs of the rule or the rule itself are merged or deleted respectively, it will result in a larger change in the fuzzy inference results and significantly affect the modeling accuracy. Thus such a rule is of greater importance. To derive a good balance between interpretability and accuracy, when MFs and rules are reduced, it is desirable to preserve all such efficient and effective rules.

Let's define the degree of the coverage of the i-th data point $\left(\mathbf{X}_i^T, \mathbf{Y}_i^T\right)$ by the k-th rule as the product of their firing strength and the membership function values as in Eq. (3-3),

$$c_{k,i} = f_k^{III} \prod_{m=1}^{n_5} \mu_{OL_k(m),m}^{IV}\left(y_{i,m}\right) \tag{3-3}$$

The *Hebbian importance* of the k-th rule is defined as the sum of the degree of coverage of all the training samples as in Eq. (3-4).

$$I_k = \sum_{i=1}^{n} c_{k,i} \tag{3-4}$$

The fuzzy rules can then be efficiently sorted by the importance of them, defined by Eq. (3-4), in a decreasing order. This is called the *Hebbian ordering* of the rules.

At the beginning of rule reduction, the original rule set, denoted by R, contains all the sorted rules and the reduced rule set, denoted by R', is null. The rules in R are presented successively according to their Hebbian orders. If there is no rule in the reduced rule set, the rule in R is directly added into R', otherwise, the fuzzy set in each dimension of the rule will be added or merged into the fuzzy sets in R' according to a set of criteria defined below. All the newly added or merged fuzzy sets will be linked together to formulate a new rule in the reduced rule set R'.

During the merging process, some fuzzy sets may be shared by several rules. To change a fuzzy set shared by many rules is equivalent to modifying these rules simultaneously, thus may exert much influence on the performance of system, while a less shared fuzzy set only affects the performance locally. Thus, much shared fuzzy set is of higher importance than the less shared ones.

Denote the importance of a fuzzy set F as $\hat{I}_F$. The $\hat{I}_F$ is a changing value during the merging process. At each time the k-th rule in the original rule set R is presented, its fuzzy sets A_i and B_j, (i.e. for the k-th rule is "IF $x_i = A_i$ THEN $y_j = B_j$") have the same importance with their associated rule. It is described in Eq. (3-5),

M1:

$$\hat{I}_{A_i} = \hat{I}_{B_j} = I_k \tag{3-5}$$

for $i = 1,...,n_1$ and $j = 1,...,n_5$.

For each input dimension i, among the fuzzy sets in the reduced rule set R' of this dimension, the fuzzy set A_i' with the maximum degree of overlap over A_i (the degree of overlap is defined in Eq. (3-9)) is selected. If the maximum degree of overlap satisfies a specified criteria, they will be merged into another fuzzy set A_i'', otherwise, the fuzzy set A_i is directly added into the reduced rule set of this dimension. The centroid and the variance of A_i and A_i', (c_{A_i}, δ_{A_i}) and $(c_{A_i'}, \delta_{A_i'})$ respectively, are merged into $(c_{A_i''}, \delta_{A_i''})$ using Eqs. (3-6) and (3-7).

M2:

$$c_{A_i''} = \frac{\hat{I}_{A_i} c_{A_i} + \hat{I}_{A_i'} c_{A_i'}}{\hat{I}_{A_i} + \hat{I}_{A_i'}} \tag{3-6}$$

$$\delta_{A_i''} = \frac{\hat{I}_{A_i} \delta_{A_i} + \hat{I}_{A_i'} \delta_{A_i'}}{\hat{I}_{A_i} + \hat{I}_{A_i'}} \tag{3-7}$$

The importance of the new fuzzy set A_i'' is given by Eq. (3-8)

M3:

$$\hat{I}_{A_i''} = \hat{I}_{A_i} + \hat{I}_{A_i'} \tag{3-8}$$

In other words, the centroid and variance of resultant fuzzy set is the weighted average of the two fuzzy sets in accordance to their importance, and the degree of importance of this final fuzzy set is the sum of the importance of the two. Then the newly generated fuzzy set A_i'' replaces the previous fuzzy set A_i'. For each output dimension, fuzzy set B_j is either added directly or merged into the fuzzy sets of this output dimension in the reduced rule set, using the same process for the input dimensions. Finally, the newly added or merged fuzzy sets in all the dimensions are linked together to formulate a fuzzy rule in the reduced rule set R'.

Given fuzzy sets A and B, the degree of overlap of A by B is defined as in Eq. (3-9).

$$S_A(B) = \left| \frac{A \cap B}{A} \right| \tag{3-9}$$

As the Gaussian membership function is used in the proposed method, the overlap measure can be derived using the centers and variances of the MFs. For the fuzzy sets A and B, with membership functions $\mu_A(x) = \exp\left\{-(x-c_A)^2/\delta_A^2\right\}$ and $\mu_B(x) = \exp\left\{-(x-c_B)/\delta_B^2\right\}$ respectively, assuming that $c_A \geq c_B$ and then $|A|$ and $|A \cap B|$ can be expressed in Eqs. (3-10) and (3-11) (Juang and Lin 1998).

$$|A| = \sqrt{\pi}\delta_A \tag{3-10}$$

$$|A \cap B| = \frac{1}{2}\left\{\frac{h^2\left[c_B - c_A + \sqrt{\pi}(\delta_A + \delta_B)\right]}{\sqrt{\pi}(\delta_A + \delta_B)} + \frac{h^2\left[c_B - c_A + \sqrt{\pi}(\delta_A - \delta_B)\right]}{\sqrt{\pi}(\delta_B - \delta_A)} + \frac{h^2\left[c_B - c_A - \sqrt{\pi}(\delta_A - \delta_B)\right]}{\sqrt{\pi}(\delta_A - \delta_B)}\right\} \tag{3-11}$$

where $h(x) = \max\{0, x\}$.

The merging criterion is that, if $S_A(B) > \lambda$ or $S_B(A) > \lambda$, they will be merged, where λ is a specified threshold that determines the maximum degree of overlap between fuzzy sets that the system can tolerate. Higher λ may increase the accuracy but would degrade the interpretability, while a lower λ may cause larger number of rules to be reduced but the risk of that the number of rules is less than necessary to maintain high modeling accuracy.

After all the rules are presented, the following steps will be executed to remove feature and delete redundant and conflicting rules in the reduced rule set R':

S1: if there is only one membership function within one dimension, this dimension (feature) will be removed;

S2: if there is any rule that has the same conditions and consequences with others, it is removed;

S3: if there are any conflicting rules that have equivalent conditions but different consequences, the one with the higher degree of importance is preserved and the others are deleted. Finally, the original rule set is replaced with the reduced rule set.

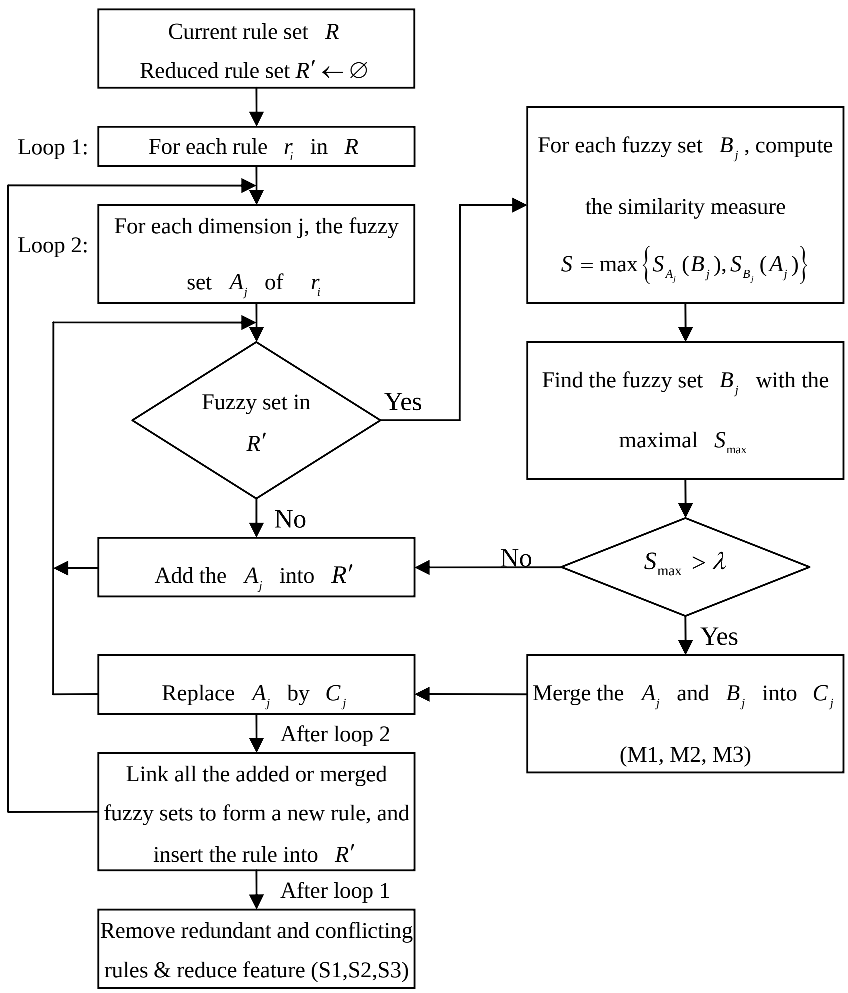
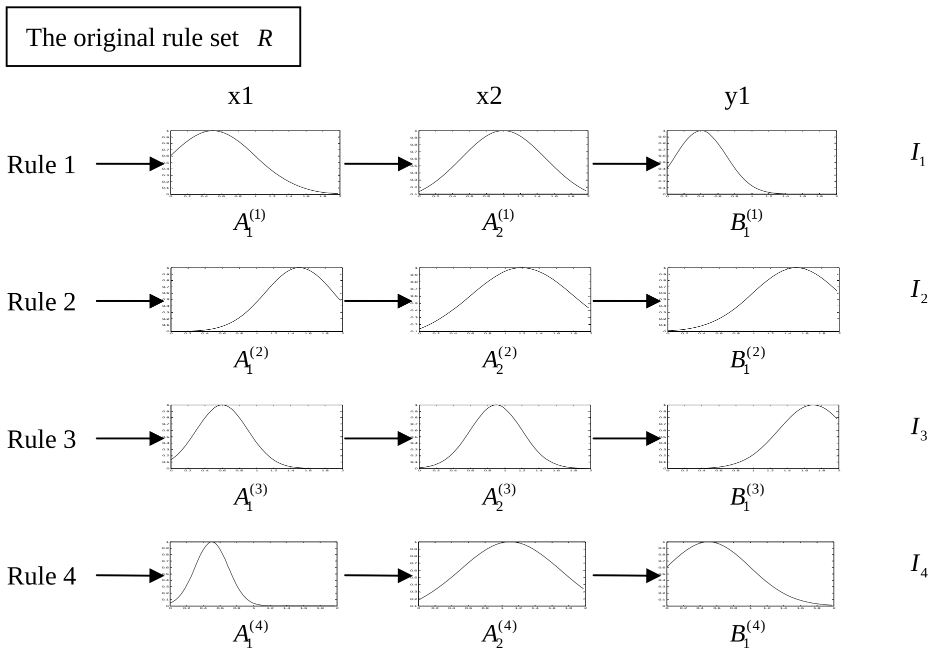


Figure 3-3: The flowchart of the rule reduction algorithm.

The flowchart of the algorithm is shown in Figure 3-3. An example is shown as follows:

There are 4 rules in the original rule set R at the beginning, shown in Figure 3-4.



M1: $\hat{I}_{A_1^{(i)}} = \hat{I}_{A_2^{(i)}} = \hat{I}_{B_1^{(i)}} = I_i$, for $i = 1, 2, 3, 4$ Hebbian ordering: $I_1 > I_2 > I_3 > I_4$

Figure 3-4: The original rule set R.

After rule 1 is presented, the 1st rule is added into the reduced rule set R' (Figure 3-5).

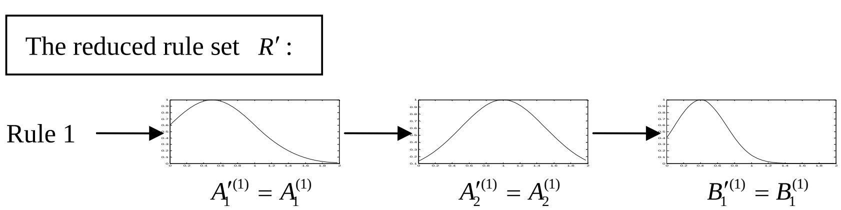


Figure 3-5: The reduced rule set R' after the presentation of rule 1.

After the 2nd rule is presented, the two fuzzy sets in the 2nd input dimension, $A_2^{(2)}$ and $A_2'^{(1)}$, are so similar that they are merged to a new $A_2'^{(1)}$. In the 1st input dimension and the output dimension, the fuzzy sets are well separated thus they are retained. They are shown in Figure 3-6.

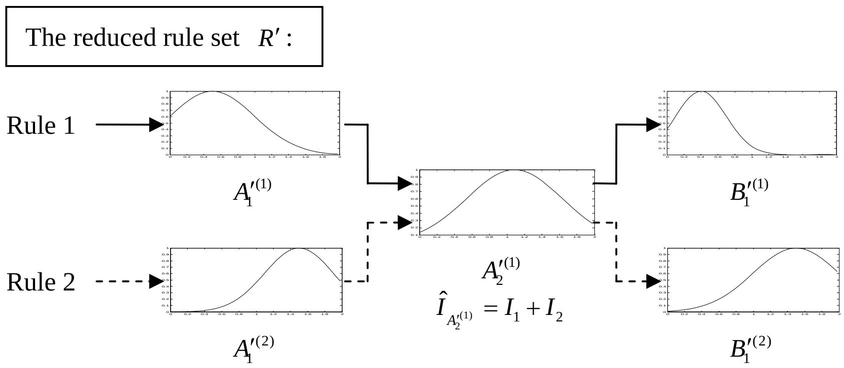


Figure 3-6: The reduced rule set R' after the presentation of rule 2.

After the 3rd rule is presented, $A_1^{(3)}$ and $A_1'^{(1)}$ are merged to a new $A_1'^{(1)}$, $A_2^{(3)}$ and $A_2'^{(1)}$ are merged to a new $A_2'^{(1)}$, and $B_1^{(3)}$ and $B_1'^{(2)}$ are merged in to a new $B_1'^{(2)}$ (See Figure 3-7).

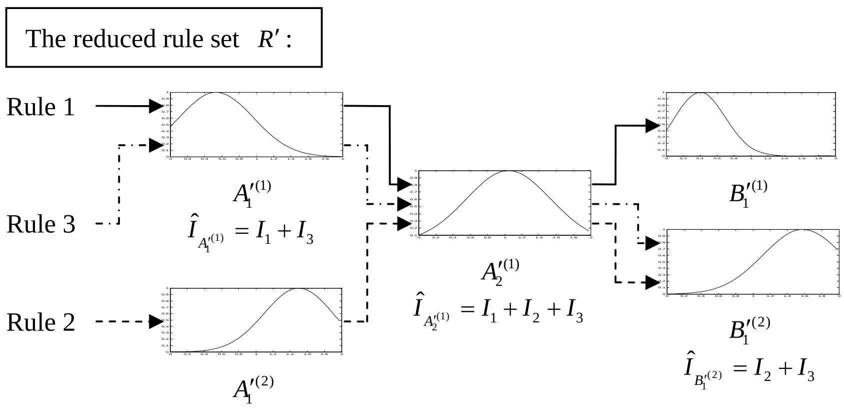


Figure 3-7: The reduced rule set R' after the presentation of rule 3.

After the 4th rule is presented, $A_1^{(4)}$ and $A_1'^{(1)}$ are merged to a new $A_1'^{(1)}$, $A_2^{(4)}$ and $A_2'^{(1)}$ are merged to a new $A_2'^{(1)}$, and $B_1^{(4)}$ and $B_1'^{(1)}$ are merged to a new $B_1'^{(1)}$ (See Figure 3-8).

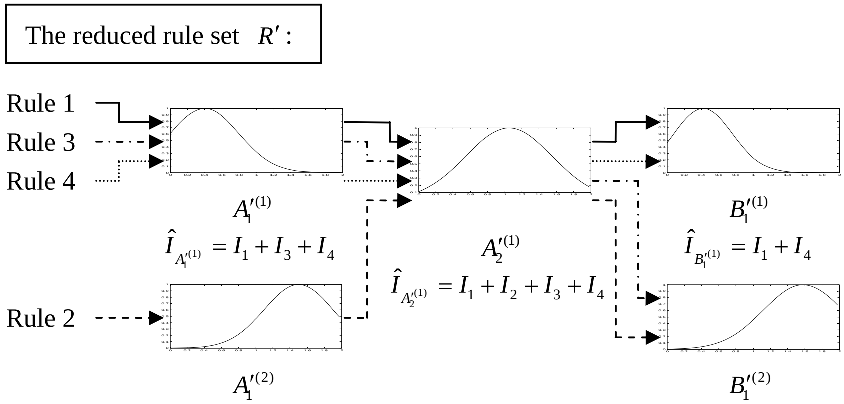


Figure 3-8: The reduced rule set R' after the presentation of the rule 4.

After all the rules are presented, we can see that, firstly, there is only one fuzzy set in the 2nd input dimension. Thus, this input feature will be removed (S1). Secondly, it is easily to see that, rule 1 and rule 4 have the same conditions and consequence. They are equivalent and one of them will be deleted (S2). In this example rule 1 is preserved as it has higher importance. Finally, rule 1 and rule 3 have the same condition but different consequence. So they are conflicting and one of them has to be removed to improve the interpretability of the system. As $I_1 > I_3$, rule 1 is preserved and rule 3 is deleted (S3). At the end, 4 rules are reduced to 2 rules and the input feature x_2 are removed.

3.1.3 Parameter tuning

To tune the parameters of the neuro-fuzzy system, the low-level learning ability of neural network is utilized. Several learning algorithms have been proposed in the neural network research domain, such as the gradient descent based methods like the Least Mean Square (LMS) algorithm and Back-Propagation (BP) algorithm (Haykin 1999), and evolutionary based methods like Genetic Algorithm (GA) (Goldberg 1989).

In this thesis, the LMS algorithm is chosen to tune the parameters of the neuro-fuzzy system, due to the following merits over other methods (Haykin 1999):

1. The LMS algorithm does not require knowledge of the statistics of the environment. It is model independent.
2. The LMS algorithm is robust. Small model uncertainty and small disturbances can only result in small estimation errors.
3. The LMS algorithm is simple and easy to implement.

The error function used is described by Eq. (3-12):

$$E = \frac{1}{2}\sum_{m=1}^{n_5}\left(o_m^V - y_m\right)^2 = \frac{1}{2}\sum_{m=1}^{n_5} e_m{}^2 \tag{3-12}$$

where $e_m = o_m^V - y_m$.

Using Eqs. (13)-(16), each parameter in the fuzzy neural network is adjusted according to the partial derivative of E to that parameter.

$$\Delta c_{oL_k(m),m}^{IV} = -\frac{\partial E}{\partial c_{OL_k(m),m}^{IV}} = -\frac{\partial E}{\partial o_m^V}\frac{\partial o_m^V}{\partial c_{OL_k(m),m}^{IV}} = -e_m \frac{o_k^{III} \Big/ \left(\delta_{OL_k(m),m}^{IV}\right)^2}{\sum_{k'=1}^{n_3} o_k^{III} \Big/ \left(\delta_{OL_{k'}(m),m}^{IV}\right)^2} \tag{3-13}$$

$$\Delta \delta_{OL_k(m),m}^{IV} = -\frac{\partial E}{\partial \delta_{OL_k(m),m}^{IV}} = -\frac{\partial E}{\partial o_m^V}\frac{\partial o_m^V}{\partial \delta_{OL_k(m),m}^{IV}}$$

$$= -2e_m \frac{o_k^{III}\left(\left(\sum_{k'=1}^{n_3} c_{OL_{k'}(m),m}^{IV} o_{k'}^{III} \Big/ \left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right) - c_{OL_k(m),m}^{IV}\left(\sum_{k'=1}^{n_3} o_{k'}^{III} \Big/ \left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right)\right)}{\left(\sum_{k'=1}^{n_3} o_{k'}^{III} \Big/ \left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right)^2 \left(\delta_{OL_k(m),m}^{IV}\right)^3} \tag{3-14}$$

$$\Delta c_{i,IL_k(i)}^{II} = -\frac{\partial E}{\partial c_{i,IL_k(i)}^{II}} = -\frac{\partial E}{\partial o_m^V}\frac{\partial o_m^V}{\partial \mu_{i,IL_k(i)}^{II}}\frac{\partial \mu_{i,IL_k(i)}^{II}}{\partial c_{i,IL_k(i)}^{II}}$$

$$= -2e_m \frac{o_k^{III}\left(c_{OL_k(m),m}^{IV}\left(\sum_{k'=1}^{n_3} o_{k'}^{III}\Big/\left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right) - \left(\sum_{k'=1}^{n_3} c_{OL_{k'}(m),m}^{IV}\, o_{k'}^{III}\Big/\left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right)\right)\left(x_i - c_{i,IL_{k'}(i)}^{II}\right)}{\left(\sum_{k'=1}^{n_3} o_{k'}^{III}\Big/\left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right)^2\left(\delta_{i,IL_k(i)}^{II}\right)^2\left(\delta_{OL_k(m),m}^{IV}\right)^2}$$

(3-15)

$$\Delta \delta_{i,IL_k(i)}^{II} = -\frac{\partial E}{\partial \delta_{i,IL_k(i)}^{II}} = -\frac{\partial E}{\partial o_m^V}\frac{\partial o_m^V}{\partial \mu_{i,IL_k(i)}^{II}}\frac{\partial \mu_{i,IL_k(i)}^{II}}{\partial \delta_{i,IL_k(i)}^{II}}$$

$$= -2e_m \frac{o_k^{III}\left(c_{OL_k(m),m}^{IV}\left(\sum_{k'=1}^{n_3} o_{k'}^{III}\Big/\left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right) - \left(\sum_{k'=1}^{n_3} c_{OL_{k'}(m),m}^{IV}\, o_{k'}^{III}\Big/\left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right)\right)\left(x_i - c_{i,IL_{k'}(i)}^{II}\right)^2}{\left(\sum_{k'=1}^{n_3} o_{k'}^{III}\Big/\left(\delta_{OL_{k'}(m),m}^{IV}\right)^2\right)^2\left(\delta_{i,IL_k(i)}^{II}\right)^3\left(\delta_{OL_k(m),m}^{IV}\right)^2}$$

(3-16)

3.2 Experimental results

There are four sets of experiments that were performed to evaluate the proposed HeRR algorithm. They are three sets of Nakanishi data set (Nakanishi, Turksen et al. 1993) (Sugeno and Yasukawa 1993), and the highway traffic flow density prediction data set.

3.2.1 Nakanishi Data Set

In this section, the proposed learning method for neuro-fuzzy system is evaluated using the benchmark experiments on three data sets: the human operation of a chemical plant, a nonlinear system and the daily price of a stock in a stock market.

The whole data sets have been published in (Nakanishi, Turksen et al. 1993). They were used to evaluate the following six reasoning methods: Mamdani version of Zadel's CRI, Turksen's version of interval-valued CRI, Turksen's versions of point-valued and interval-valued AARS, and Sugeno's version of position and gradient type of reasoning (Nakanishi, Turksen et al. 1993). These data sets are also used to evaluate the qualitative modeling based on fuzzy logic (Sugeno and Yasukawa 1993). Each of the three data sets is split into three groups: A, B, and C (Nakanishi, Turksen et al. 1993). In the following experiments, the data sets of groups A and B are used as the training sets while the group C is used as the testing set. The experimental results are compared against 1) the Pseudo Outer-Product-based Fuzzy Neural Network (POPFNN) (Ang, Quek et al. 2003); 2) and its modified version: Rough-Set-based Pseudo Outer-Product fuzzy rule identification algorithm (RSPOP) (Ang and Quek 2005); 3) Adaptive-Network-based Fuzzy Inference System (ANFIS) (Jang 1993); 4) Evolving Fuzzy Neural Networks (EFuNN) (Kasabov 2001); 5) and Dynamic Evolving Neural Fuzzy Inference System (DENFIS) (Kasabov and Song 2002). To evaluate the performance of these systems, both the Mean Squared Error (MSE) and the correlation coefficient (R) are used in the experiments. The comparative analysis is presented in Section 3.2.5.

3.2.1.1 A Nonlinear System

The non-linear system is described in Eq. (3-17).

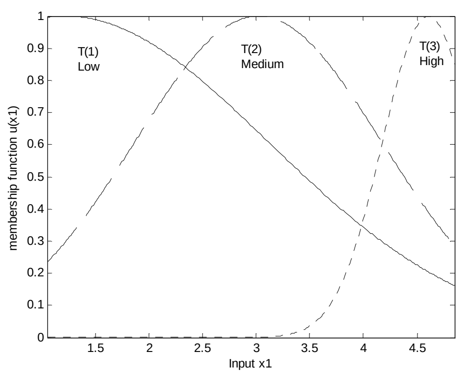
$$y = \left(1 + x_1^{-2} + x_2^{1.5}\right)^2 \qquad 1 \le x_1, x_5 \le 5 \tag{3-17}$$

The data set consists of 4 input variables and 1 output variable, where only x_1 and x_2 are useful in the modeling while x_3 and x_4 are irrelevant variables.

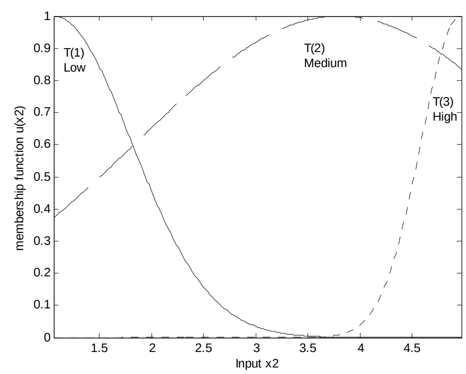
Figure 3-9 and Table 3-1 show the shapes, the centroids and the widths of the derived Gaussian membership functions. For this data set, the feature selection method in (Nakanishi, Turksen et al. 1993) discards the input features x_3 and x_4, while the RSPOP (Ang and Quek 2005) discards the feature x_3 and partially discards x_4. In our proposed method, the two irrelevant variables x_3 and x_4 are discarded too. There are 3 fuzzy sets for each of the input features x_1, x_2 and output variable y_1 (see Figure 3-9). In total, 7 rules are generated, as shown in Table 3-2. To show the interpretability, the first two rules are taken for example. The 1st rule indicates that, if x_1 takes low value and x_2 takes medium value, then y_1 will be a medium value, and the 2nd rule indicates that, if x_1 is high and x_2 is medium, then y_1 will be low. From Eq. (3-17), we can see that, for x_1, higher value decreases y and lower value increases y, while for x_2, the higher the value, the higher the value for y, and vice versa. From these two rules, it can be observed that, while the x_2 are taking the same fuzzy label, the output y_1 is inversely correlated to x_1. It is quite understandable to human.

Label		Low	Medium	High
x_1	c_1^{II}	1.219	3.020	4.605
	δ_1^{II}	1.900	1.155	0.428
x_2	c_2^{II}	1.061	3.803	4.962
	δ_2^{II}	0.747	1.953	0.379
y	c^{IV}	1.337	3.321	5.050
	δ^{IV}	0.816	1.875	1.873

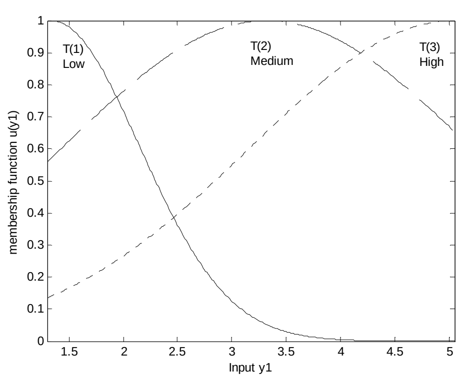
Table 3-1: The membership functions for the non-linear system modeling problem.



(a) Input x_1



(b) Input x_2



(c) Output y

Figure 3-9: Membership functions for the non-linear system modeling problem.

Rule	x_1	x_2	y
1	Low	Medium	Medium
2	High	Medium	Low
3	Low	High	High
4	Medium	Medium	Low
5	Medium	High	Medium
6	Medium	Low	Low
7	Low	Low	Low

Table 3-2: Fuzzy rules for the non-linear system modeling problem.

3.2.1.2 Human Operation of a Chemical Plant

The problem is to model the human operation of a chemical plant. There are 5 input variables and 1 output variable, as shown in Table 3-3.

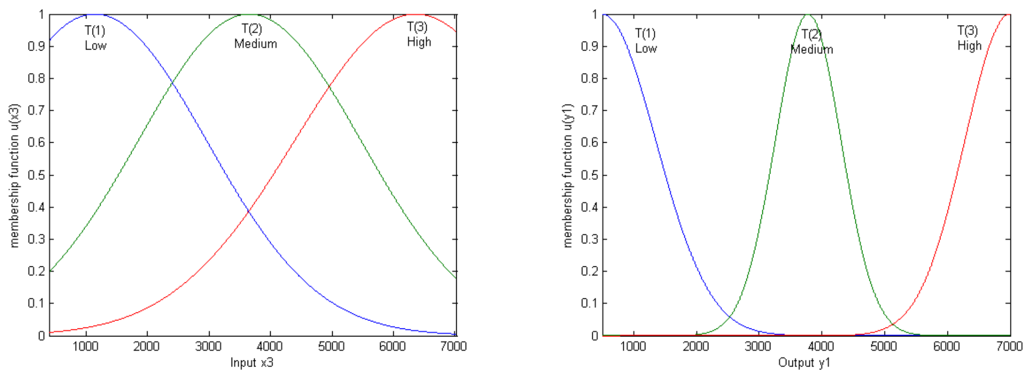
x_1	Monomer concentration
x_2	Change of monomer concentration
x_3	Monomer flow rate
x_4	Local temperatures inside the plant
x_5	Local temperatures inside the plant
y	Set point for monomer flow rate

Table 3-3: The variables in the data set of the human operation of a chemical plant.

The feature selection method in Nakanishi et al. (1993) discards the inputs x_2, x_4 and x_5, whereas RSPOP discards the inputs x_1, x_2 and x_5. In our proposed method, among the five input dimensions, x_1 to x_5, only x_3 is preserved and others are discarded. There are 3 fuzzy sets in both input x_3 and the output y. The shapes, the centroids and the widths of the derived Gaussian membership functions are shown in Figure 3-10 and Table 3-4. Three rules generated from the system, shown in Table 3-5. From the rules, it can be see that, the output y is positively correlated with the input x_3.

Label		Low	Medium	High
x_3	c^{II}	1153.039	3660.854	6365.789
	δ^{II}	1808.950	1810.971	1972.284
y	c^{IV}	500.135	3781.490	7000.000
	δ^{IV}	854.602	524.820	721.802

Table 3-4: The centroid and width of the membership functions for the modeling problem of the human operation of a chemical plant.



(a) Input x_3

(b) Output y

Figure 3-10: Membership functions for the modeling problem of the human operation of a chemical plant.

Rule	x_3	y
1	Low	Low
2	Medium	Medium
3	High	High

Table 3-5: Fuzzy rules for the modeling problem of the human operation of a chemical plant.

3.2.1.3 Daily Price of a Stock in a Stock Market

This problem is to predict the stock price using various features of a stock in a stock market. There are 10 input variables and 1 output variable in the data set. They are summarized in Table 3-6.

x_1	Past change of moving average over a middle period
x_2	Present change of moving average over a middle period
x_3	Past separation ratio with respect to moving average over a middle period
x_4	Present separation ratio with respect to moving average over a middle period
x_5	Present change of moving over a short period
x_6	Past change of price, for instance, change on one day before
x_7	Present change of price
x_8	Past separation ratio with respect to moving average, over a short period
x_9	Present change of moving average over a long period
x_{10}	Present separation ratio with respect to moving average over a short period
y	Prediction of stock price

Table 3-6: The variables in the data set of daily stock price prediction.

The proposed method discards the input variable x_2 and x_5, while the feature selection method in Nakanishi et al. (1993) discards the inputs x_1, x_2, x_3, x_6, x_7, x_9, and x_{10}, and the RSPOP discards the inputs x_1, x_2, x_3, x_6, and x_{10}. Figure 3-11 and Table 3-7 show the shapes, the centroids and the widths of the derived Gaussian membership functions. There are in total 20 fuzzy rules identified by the proposed method, shown in Table 3-8.

Label		Low	Medium	High
x_1	c_1^{II}	0.023	N	35.000
	δ_1^{II}	3.509	N	3.511
x_3	c_3^{II}	-10.217	10.127	24.767
	δ_3^{II}	4.624	7.421	7.724
x_4	c_4^{II}	-10.315	6.596	22.144
	δ_4^{II}	5.241	7.057	15.891
x_6	c_6^{II}	-14.300	N	9.770
	δ_6^{II}	2.450	N	12.250
x_7	c_7^{II}	-8.882	N	6.814
	δ_7^{II}	2.445	N	10.879
x_8	c_8^{II}	-19.690	-12.837	9.936
	δ_8^{II}	3.161	3.166	12.164
x_9	c_9^{II}	-18.930	-12.208	6.655
	δ_9^{II}	3.460	2.624	11.237
x_{10}	c_{10}^{II}	-19.690	-7.306	8.429
	δ_{10}^{II}	3.161	8.675	7.194
y	c^{IV}	-21.209	3.765	34.510
	δ^{IV}	7.273	7.911	7.337

Table 3-7: The centroid and width of the membership functions for the daily stock price prediction problem

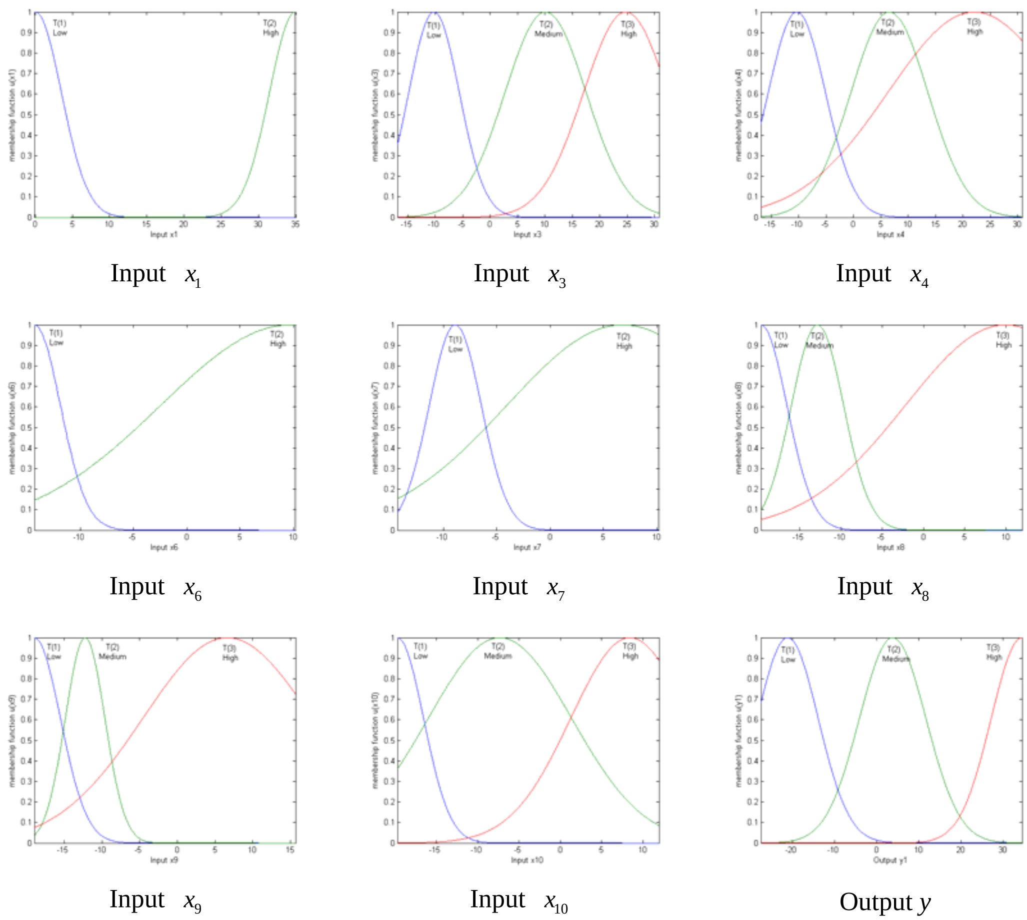


Figure 3-11: Membership functions for the daily stock price prediction problem.

Rule	x_1	x_3	x_4	x_6	x_7	x_8	x_9	x_{10}	y
1	Low	High	High	High	High	High	High	High	Low
2	Low	Medium	High	High	High	High	High	High	Low
3	Low	High	Low	High	High	High	High	High	Medium
4	Low	Low	Low	High	High	High	High	High	Medium
5	Low	Low	Low	High	High	High	High	Medium	Medium
6	Low	Medium	High	High	High	High	High	Medium	Low
7	Low	High	Low	High	High	High	High	Medium	Medium
8	High	Medium	Low	High	High	Medium	High	Medium	Medium
9	Low	Medium	Low	Low	High	High	Low	Low	Medium
10	Low	Low	Medium	High	High	High	High	High	Medium
11	Low	Medium	Low	High	High	Medium	High	Medium	Medium
12	Low	Medium	Low	High	High	High	High	Medium	Medium
13	Low	High	Low	High	High	Medium	High	High	Medium
14	High	Low	Medium	High	High	High	High	Medium	Medium
15	Low	Medium	Medium	High	High	High	High	Medium	High
16	Low	Medium	Medium	High	High	High	High	High	Medium
17	High	High	Medium	High	High	High	High	Medium	High
18	Low	Medium	Medium	High	High	Low	Medium	Medium	Medium
19	Low	Medium	Medium	High	High	High	Medium	Medium	Low
20	Low	Medium	Low	High	Low	High	High	Medium	Medium

Table 3-8: The fuzzy rules for the daily stock price prediction problem.

3.2.1.4 Discussion

Table 3-9 shows the consolidated experimental results based on the Nakanishi dataset in terms of the Mean Squared Error (MSE) and the Pearson's correlation coefficient (R). On the problem of modeling the human operation of a chemical plant, although only one variable is reserved and three fuzzy rules are generated by the proposed method, the modeling error (MSE) and accuracy (R) are the best among all the 9 models. Good interpretability and accuracy are achieved simultaneously (see Tables 3-9 and 3-10). On the modeling of a non-linear system, the proposed method correctly identifies the useful variables. It outperforms others for the two measures. On the prediction of stock price, as described before, the inputs x_1, x_2, x_3, x_6, x_7, x_9, and x_{10} are discarded by the feature selection method in Nakanishi et al. (1993), and the inputs x_1, x_2, x_3, x_6, and x_{10} are discarded by RSPOP. In the proposed model, x_2 and x_5 are discarded. Input x_2 is the one that has been discarded by all of them but neither of them considered input x_5 as a redundant feature. From the results, although the proposed model uses a larger set of attributes, it has achieved superior performance to the other models. It indicates that some variables discarded by other models are still useful in the prediction and cannot be judiciously discarded.

As both the RSPOP model and the proposed HeRR model employ Hebbian-like learning method and reduce the rule set and attribute set simultaneously with the aim to pursue a balance between accuracy and interpretability, they are contrasted in Table 3-10. The proposed model begins with a rule set that is larger than RSPOP, for the first two data sets, and an equal size for the third one. After the tuning and pruning, the HeRR ends up with a much smaller size of rule set on for all of the 3 data sets than that produced by RSPOP. The ratio of the decrease of the number of the rules by the proposed method is larger than that by RSPOP. For the MSE, the HeRR outperforms RSPOP with the ratio of improvement of 88.6%, 51.7% and 39.1% for the three data sets respectively. For R, the ratio of improvement by the proposed method are 1.5%, 6.0% and 2.6% for the three data sets respectively. It is safe to conclude that the proposed HeRR achieves better performance than RSPOP on both interpretability and accuracy.

Data set	Chemical plant		Nonlinear system		Stock prediction	
	MSE	R	MSE	R	MSE	R
HeRR	2.423×10^4	0.998	0.185	0.911	15.138	0.947
POPFNN	5.630×10^5	0.946	0.270	0.877	76.221	0.733
RSPOP	2.124×10^5	0.983	0.383	0.856	24.859	0.922
Sugeno (P-G)	1.931×10^6	0.990	0.467	0.845	168.90	0.700
Mandani	6.580×10^5	0.937	0.862	0.490	40.84	0.865
Turksen (IVCRI)	2.581×10^5	0.993	0.706	0.609	93.02	0.661
ANFIS	2.968×10^6	0.780	0.286	0.853	38.062	0.875
EFuNN	7.247×10^5	0.946	0.566	0.720	72.542	0.756
DENFIS	5.240×10^4	0.995	0.411	0.805	69.824	0.810

Table 3-9: Consolidated experimental results on Nakanishi Data sets.

Data set		Chemical plant	Nonlinear system	Stock Prediction
HeRR	No. of rules before reduction	30	25	50
	No. of rules after reduction	3	7	20
	Rule Reduction (%)	90.0	64.0	60.0
	MSE	2.423×10^4	0.185	15.138
	R	0.998	0.911	0.947
RSPOP	No. of rules before reduction	24	22	50
	No. of rules after reduction	14	17	29
	Rule Reduction (%)	41.7	22.7	14.0
	MSE	2.124×10^5	0.383	24.859
	R	0.983	0.856	0.922
Improvement (HeRR against RSPOP)	Rule Reduction (%)	48.3	41.3	46
	MSE (%)	88.6	51.7	39.1
	R (%)	1.5	6.0	2.6

Table 3-10: Benchmark against RSPOP

To evaluate its performance on noisy data, Gaussian noise are added into the three Nakanishi data sets, with 5%, 10% and 15% noise level. The results on noisy data, compared with the noise-free data, are shown in Table 3-11. From the table we can see that, the accuracy decreases as the noise level increases. But the performance does not degrade very much. Referring to Table 3-9, although noises appear, the accuracy is still better than most of the other benchmarked systems. It indicates that the system is tolerant to noises in data.

Data set	Chemical plant		Nonlinear system		Stock prediction	
	MSE	R	MSE	R	MSE	R
Noise-free	2.423×10^{4}	0.998	0.185	0.911	15.138	0.947
5% noise	2.575×10^{4}	0.997	0.318	0.843	17.019	0.940
10% noise	3.989×10^{4}	0.995	0.336	0.825	22.608	0.919
15% noise	5.188×10^{4}	0.993	0.407	0.798	25.994	0.909

Table 3-11: Comparison on noise-free and noisy data with 5%, 10% and 15% noise levels.

The parameter settings of the above results are shown in Table A-1, in appendix A.

3.2.2 Highway Traffic Flow Density Prediction

This experiment attempts to evaluate the performance of the proposed HeRR learning method on the highway traffic flow density data set. The raw traffic flow data for the benchmark test was obtained from (Tan 1997). The raw data were collected at site 29 located at exit 15 along the eastbound Pan Island Expressway (PIE) in Singapore using loop detectors embedded beneath the road surface. There are a total of five lanes at the site, two exit lanes and three straight lanes for the main traffic. In this experiment, only the traffic flow data for the three straight lanes are employed. There are four input attributes in this data set, the time and the traffic density of the three lanes. The traffic flow trend at the site are modeled by the proposed HeRR neuro-fuzzy model and the trained network is used to predict traffic density for time $t+\tau$, where $\tau=5$, 10, 30, 45 and 60 minutes. Figure 3-12 shows the plotting of the traffic flow density data for the three straight lanes spanning a period of six days from 5th to 10th September 1996. The whole data set is divided into 3 cross-validation groups of training and testing sets, namely CV1, CV2 and CV3, which are the same with that in (Tung and Quek 2002) and (Ang and Quek 2005). The Pearson correlation coefficient (R) and the Mean Squared Error (MSE) are used to evaluate the performance. The proposed method is compared against RSPOP, POPFNN, GenSoFNN (Tung and Quek 2002), EFuNN and DENFIS. The consolidated experimental results are shown in Figure 3-13. Figure 3-13 (a), (c) and (e) show the average R of the three cross-validation groups, whereas Figure 3-13 (b), (d) and (f) show the average MSE of them. The proposed HeRR method shows superior performance to the other models, where it achieves the highest average R and the lowest average MSE of lane 1 to 3 for all the intervals τ =5, 15, 30, 45 and 60. Table 3-12 shows the average R, MSE and the number of rules identified, for all the time intervals of the three lanes. The proposed HeRR method derives on the average only 8.1 fuzzy rules, which is lowest among all of these models, while its modeling accuracy is superior. It indicates that, the proposed method yields a good balance of accuracy and interpretability.

The parameter settings of the results are shown in Tables A-2 to A-4, in appendix A.

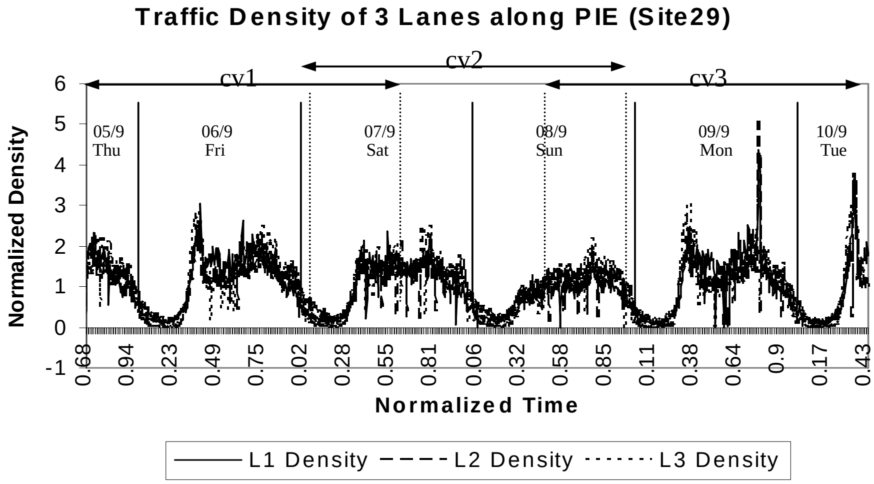
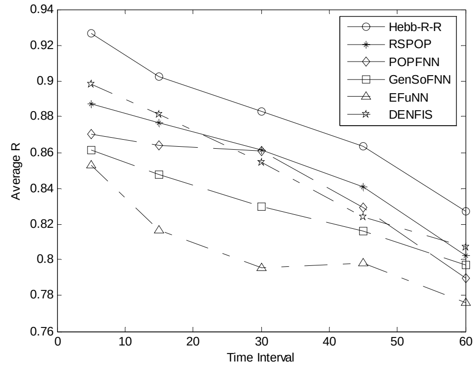


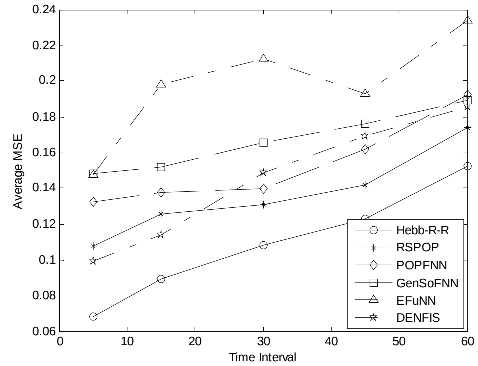
Figure 3-12: The plotting of the traffic flow density and the division of the cross-validation groups.

Neuro-fuzzy models	Average of MSE	Average of R	Average of number of rules
HeRR	0.114	0.864	8.1
RSPOP	0.146	0.834	14.4
POPFNN	0.173	0.814	40.0
GenSoFNN	0.164	0.813	50.0
EFuNN	0.189	0.798	234.5
DENFIS	0.153	0.831	9.7

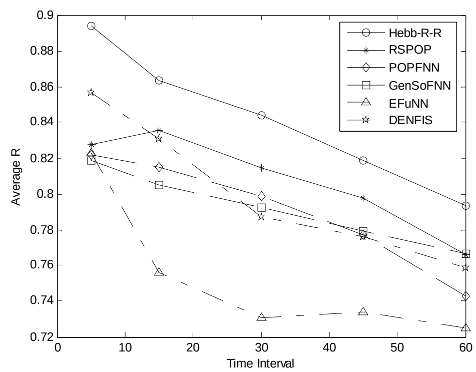
Table 3-12: Average MSE, R and number of rules.



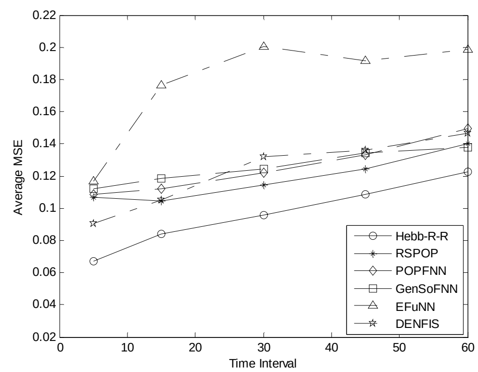
(a)



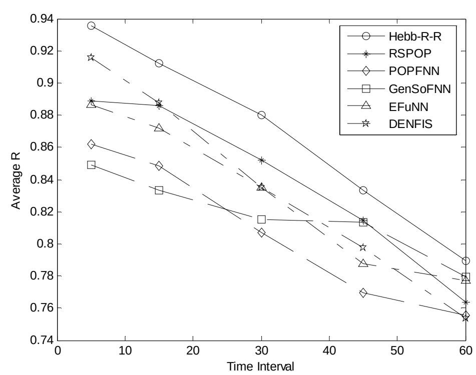
(b)



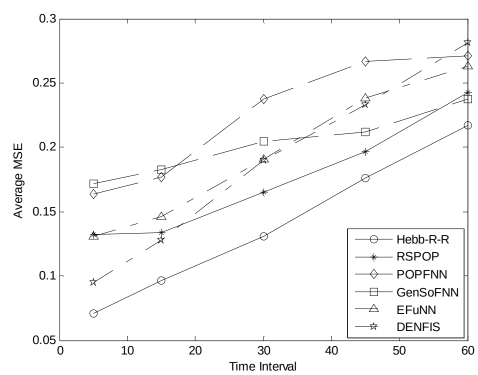
(c)



(d)



(e)



(f)

Figure 3-13: The average R and MSE of the prediction in the 3 cross-validation groups.

3.3 Summary

This chapter describes a novel method for balancing the interpretability and the demand of modeling accuracy in a neuro-fuzzy system. The interpretability is improved through a judicious reduction of the membership functions, fuzzy rules and attributes, while the accuracy is boosted through the Least Mean Square (LMS) learning algorithm that tunes the rules and MFs to the most appropriate locations in the feature space. The Hebbian ordering is proposed to represent the importance of each rule. The rule with a higher Hebbian ordering has a greater degree of coverage of the sample points and more contribution towards the modeling of the data, and is more likely to be preserved. Membership functions are merged according to the Hebbian importance of their associated rule. The resultant equivalent rules are deleted and the attributes with only one MF are removed. The problem of rule confliction is resolved by retaining the rules of higher importance. The proposed membership function merger process does not only reduce similar membership functions, but also preserves the more informative rules (the rules in high Hebbian ordering) to make the consequent LMS learning process achieve a superior accuracy.

The key strengths of the proposed method are:

1) An intuitive, consistent and compact rule base can be extracted from the trained neuro-fuzzy model, with a reduced set of attributes, clearly separating the membership functions in each dimension, and most importantly a small number of judiciously derived rules.

2) The neuro-fuzzy model does not need any clustering method like LVQ (Kohonen 1982) to specify the number and positions of the membership functions in each dimension prior to the identification of rules, unlike that in (Tung and Quek 2002; Ang, Quek et al. 2003; Ang and Quek 2005)

3) Good balance between interpretability and accuracy can be obtained through the proposed iterative processing by HeRR and LMS learning algorithm.

The performance of the proposed HeRR method is evaluated using five data sets: 1) human operation of a chemical plant; 2) a nonlinear system; 3) daily price of a stock in a stock market; 4) highway traffic flow density prediction. The experimental results show consistently good performance by the proposed method, when benchmarking against other established neuro-fuzzy models.

In the next chapter, the proposed HeRR method will be generalized into the pattern classification problems. The rough set will be hybridized with the neuro-fuzzy system for knowledge reduction.

CHAPTER

4

4 RS-HeRR: A Rough-Set based Hebbian Rule Reduction Neuro-Fuzzy System

"Our brightest blazes are
commonly kindled by unexpected sparks."
- Samuel Johnson

In previous Chapter, the proposal of the HeRR neuro-fuzzy system is primarily for the regression problems, in which the outputs represent the values of continuous variables. In this Chapter, the HeRR neuro-fuzzy system is generalized to the pattern classification problems, which involve the task that assigns the inputs to one of a number of discrete classes or categories. The Rough-Set based Hebbian Rule Reduction (RS-HeRR) neuro-fuzzy system is proposed for this task. Section 4.1.1 describes the fuzzy classifier. Section 4.1.2 presents the HeRR neuro-fuzzy classifier, which is modified from the HeRR neuro-fuzzy system (see Chapter 3). Section 4.2 proposes the hybridization of the HeRR neuro-fuzzy classifier and rough set. A rough-set based attribute reduction algorithm is proposed for knowledge reduction in fuzzy rule set. Two sets of experiments are conducted: the first set (see Section 4.3.1), including three low-dimensional benchmark experiments, is used to evaluate the classification capability of the HeRR neuro-fuzzy classifier; the second set (see

Section 4.3.2), including five real-world data sets, are used to evaluate the RS-HeRR neuro-fuzzy classifier.

4.1 Hebbian based rule reduction neuro-fuzzy system for pattern classification problem

4.1.1 Fuzzy classifier

Fuzzy classifier is defined as a set of fuzzy classification rules. Let the input vector be $\mathbf{X}^{\mathrm{T}} = [x_1, x_2, \ldots, x_i, \ldots, x_{n1}]$ (n_1 is the input dimension), and its class label be y. An example of fuzzy classification rule is in the following form (Figure 4-1):

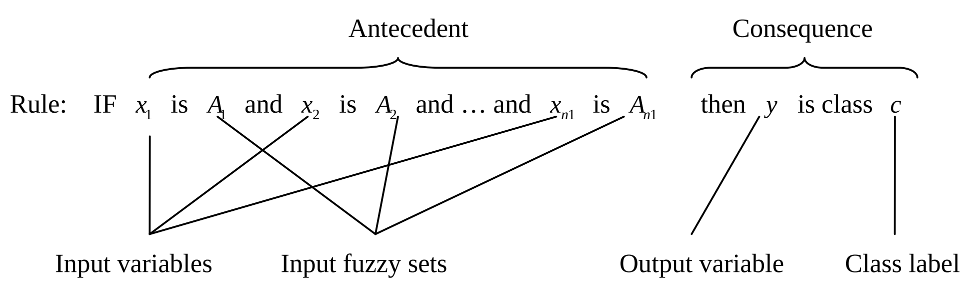


Figure 4-1: An example of fuzzy classification rule.

As in Chapter 3, the Gaussian membership function is used in the input fuzzy sets. The centroids and widths of the MFs are denoted as $(c_{i,j}^{II}, \delta_{i,j}^{II})$ for the i-th input dimension and the j-th MF. The input label in the i-th input dimension of the k-th rule is denoted $iL_k(i)$. The output label is denoted oL_k, which may take a class c_k, of the k-th rule, where $c_k = 1, 2, ..., c_{\max}$ and $c_{\max}$ is the maximal number of the class labels.

The minimum and maximum operations are employed as the fuzzy union and intersection operators respectively; thus, it is a *min-max fuzzy classifier* (Abe and Lan 1995). The computation of the fuzzy classifier can be expressed in Eq. (4-1).

$$o = c_{k'}, \quad k' = \arg\max_{k}\left(f_k\right) \tag{4-1}$$

where $f_k = \min_i\left(\exp\left(\frac{\left(x_i - c^{II}_{i,iL_k(i)}\right)^2}{\left(\delta^{II}_{i,iL_k(i)}\right)^2}\right)\right)$ is called the firing strength of the input point $\mathbf{X}^{\mathrm{T}}$ by the k-th rule.

4.1.2 Hebbian based rule reduction algorithm for classification problem

To reformulate HeRR as a classifier, the following modifications are made:

(1) The reduction phase of HeRR is performed, without the tuning phase. Through the experiments on plentiful data sets, it is shown that, the LMS algorithm, which is used for tuning the membership functions, is able to improve the modeling accuracy effectively for regression problems. However, for classification problems, its performance is unsatisfactory. A possible reason is that the output function of the system has been reformulated using min and max operators, described in Eq. (4-1). Whenever the data comes, the rule with the maximal firing strength is selected for classification. Thus, the tuning is conducted for the membership functions of that rule. However, these MFs may be shared by several rules. The tuning process will result in the changes of other classification rules,

which may deteriorate the overall classification accuracy. In this Chapter, the reduction part is preserved to maintain the interpretability of the derived rules. The balance of interpretability and accuracy will be handled in the next section.

(2) The Hebbian importance is redefined. Since the output function of the neuro-fuzzy system has been changed from Eq. (3-2) to Eq. (4-1) for classification problems, the Hebbian importance defined in Eqs. (3-3) and (3-4), is modified as follows: the Hebbian importance of the k-th rule is defined as in Eqs. (4-2) and (4-3).

$$I_k = \sum_{i=1}^{n} f_{k,i} \times \mu_{OL_k}(y_i) \tag{4-2}$$

where $f_{k,i}$ is the firing strength of the k-th rule on the i-th data point and

$$\mu_{OL_k}(y_i) = \begin{cases} 0 & \text{if } c_k \neq y_i \\ 1 & \text{if } c_k = y_i \end{cases} \tag{4-3}$$

The remaining parts of HeRR is the same as that in Chapter 3, including the initial rule generation, rule ranking, MF merging and redundancy removing (see Sections 3.1.1 and 3.1.2).

4.2 Rough-set based attribute reduction algorithm

In neuro-fuzzy systems, the knowledge is represented as a set of attributes and fuzzy rules. The reduction of ineffective attributes and redundant rules is essential to the improvement of interpretability in neuro-fuzzy systems, particularly for high-dimensional problems. Furthermore, the removal of redundancy does not only speed up the fuzzy inference process, but may also enhance classification accuracy.

In the last Chapter, the reduction of redundant attributes and rules has been done to some extent when merging the MFs (see Section 3.1.2). However, this reduction is quite weak. That is because, it only reduces the attribute which has only one fuzzy set left in that dimension, but has not considered the contributions of the attributes to the rule set and the classification accuracy. For example, if a conditional attribute, which is unrelated to the decision attribute, cannot be merged into a membership function, this conditional attribute will become redundant for the rule set. Hence it is desirable to perform the reduction of redundancy from the rule set further.

In this Chapter, a rough-set based attribute reduction algorithm is proposed to remove redundancy from the fuzzy rule set. The hybridization of the HeRR classifier and rough set, namely RS-HeRR is proposed for the pattern classification problems.

There have been some rough-set based methods to remove attributes and prune rules for fuzzy system. The QuickReduct algorithm (Shen and Chouchoulas 2002) are proposed to derive a minimal reduct of the attribute set. The QuickReduct uses a forward selection scheme, which starts with an empty attribute set and adds the attribute with maximal partial dependency among the rest of the attributes. It uses the partial dependency as a guide to search for a reduct. However, as they only consider partial dependencies, it does not guarantee that the classification performance is adversely affected after the selection. For this reason, the RSPOP (Ang and Quek 2005) proposes a method that utilize the system's performance to guide the selection process. It removes the attributes that does not decrease the accuracy, until any removal will result in inconsistency of the rule set. However, there are many times that several attributes are all satisfied. This method simply chooses any of them. Though the removal of the chosen attribute does not decrease the accuracy, it may be closely related to the decision attribute and should be preserved. Based on these observations, this thesis proposed a rough-set based attribute set selection algorithm that takes the advantages of the research work mentioned above. It uses both the partial dependency and the system's performance to select attributes. The detailed description will be given later.

The hybridization of HeRR and rough set is depicted in Figure 4-2. Firstly, whenever training samples are available, an initial rule set is generated by the HeRR classifier. Second, the proposed rough-set based attribute reduction algorithm is used as a post-processing step to simplify the rule set further. It uses the classification accuracy of the induced rule set to evaluate the selected attributes, and terminates without deteriorating the overall classification accuracy.

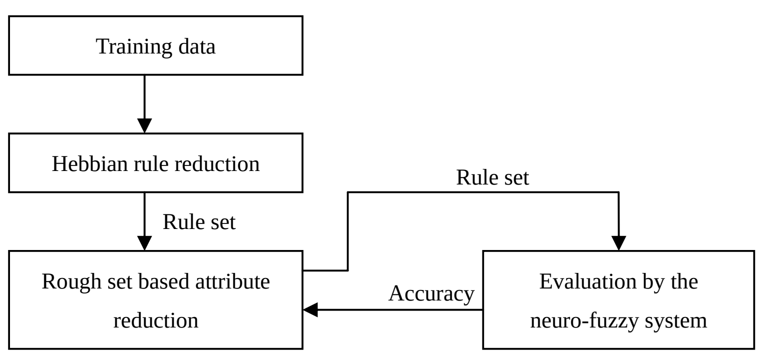


Figure 4-2: The hybrid of HeRR neuro-fuzzy system and the rough set.

The pseudo-code of the attribute selection algorithm is summarized in Figure 4-3. The algorithm begins with an empty set of attributes. The next attribute to be added is the attribute that results in the largest increase in the classification performance (e.g. the number of correctly classified samples) of the training set using the neuro-fuzzy system. If there are several attributes that give the same quantity of improvement in classification rate, the one that contributes the most to the partial dependency (defined in Eq. (2-10)) is selected. If the tie occurs again, any one is suitable. The algorithm terminates when the partial dependency becomes one and the classification accuracy using the subset of attributes is not worse than that using the whole set of the attributes.

The algorithm conducts the knowledge reduction to improve the interpretability of the rule set. The purpose of employing the classification performance to evaluate the attributes is to strike a balance between interpretability and accuracy. At the end of the algorithm, the classification accuracy using the attribute subset will be no worse than that using the whole set of attributes. Hence, it improves the interpretability without affecting the accuracy.

Algorithm

Input: $RuleSet(A,C)$, the rule set formed by the condition attribute set A and class attribute C.

Output: R, the attribute subset, $R \subseteq A$.

Denote: $ACCU(RuleSet(A,C))$, the classification accuracy of the fuzzy system using the rule set formed by attribute sets A and C.

(1) $O_A \leftarrow ACCU(RuleSet(A,C))$

(2) $R \leftarrow \varnothing$

(3) do

(4) $\quad T \leftarrow R$

(5) $\quad$ for each $a \in (A-R)$

(6) $\quad\quad O_{R\cup\{a\}} = ACCU(RuleSet(R\cup\{a\},C))$

(7) $\quad\quad$ if $O_{R\cup\{a\}} > O_T$ or $(O_{R\cup\{a\}} = O_T$ and $\gamma_{R\cup\{a\}}(C) > \gamma_T(C))$ then

(8) $\quad\quad\quad T \leftarrow R\cup\{a\}$

(9) $\quad R \leftarrow T$

(10) until $O_R \geq O_A$ and $\gamma_R(C) = 1$

(11) return R

Figure 4-3: The pseudo-code of the rough-set based attribute selection algorithm.

4.3 Experimental results and analysis

Two sets of experiments are presented in the section. The first one consists of three low-dimensional datasets to evaluate the HeRR classifier, without the rough set based attribute reduction process. The second one contains five high-dimensional datasets to evaluate the proposed RS-HeRR neuro-fuzzy system.

4.3.1 Experimental results on HeRR for low-dimensional classification problems

In this section, three experiments are used to evaluate the Hebbian based rule reduction neuro-fuzzy system for classification problem. They are (1) the *Pat* synthetic pattern classification, (2) the Two-Spiral classification, and (3) the Iris classification.

4.3.1.1 *Pat* Synthetic Pattern Classification

The *Pat* synthetic data set consists of 880 two-dimensional pattern points, depicted in Figure 4-4. There are three linearly non-separable classes. The figure is marked with class 1(c_1) and 2(c_2), while class 3(c_3) corresponds to the background region. In this experiment, 10 percent of the data are used as the training set while the remaining data are used as the testing set. The Modular Rough-Fuzzy MLP (Mitra, Mitra et al. 2001; Pal, Mitra et al. 2003), EFuNN, RSPOP (Ang and Quek 2005), and GenSoFNN (Tung and Quek 2002) are compared with the proposed HeRR system. The testing accuracy and the number of derived rules are used to evaluate the performance of the systems.

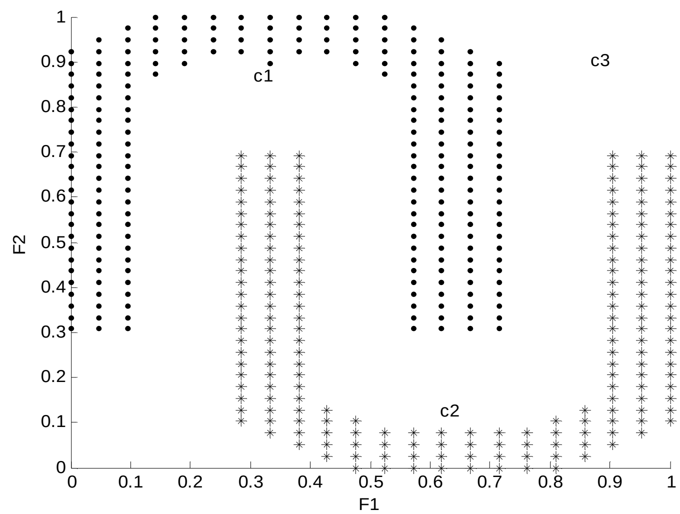


Figure 4-4: Synthetic dataset *Pat*.

Models	Accuracy (%)	No. of rule
HeRR	84.24	28
Modular rough-fuzzy MLP	70.31	8
EFuNN	82.25	88
RSPOP	55.81	16
GenSoFNN	31.19	28

Table 4-1: Comparison of performance of the different models.

The performance of different models is shown in the Table 4-1. The proposed HeRR outperforms the other benchmarked models. Its accuracy is the highest among the models, which indicates its good capability for classification task. However, to achieve such superior performance the proposed HeRR needs more rules than the Rough-Fuzzy MLP. The accuracy of EFuNN is near to that of HeRR, but it requires much more rules for classification. For the RSPOP and GenSoFNN, their performance is quite worse than the other three systems.

4.3.1.2 Two-Spiral classification

The Two-Spiral classification problem is a benchmark task for neural networks (Lang and Witbrock 1988). The task involves classifying correctly the points of two intertwined spirals. The spiral data are shown in Figure 4-5, where the two classes denote the two spirals respectively. The training set is the standard spiral data containing 194 samples, with 97 samples from each class. The testing set consists of two dense spirals with 385 samples from each class and is generated using Eqs. (4-4) - (4-6).

$$\text{spiral 1: } \begin{cases} x = \gamma \cos\theta \\ y = \gamma \sin\theta \end{cases} \tag{4-4}$$

$$\text{spiral 2: } \begin{cases} x = -\gamma \cos\theta \\ y = -\gamma \sin\theta \end{cases} \tag{4-5}$$

where

$$\gamma = \frac{1}{\pi}\left(\theta + \frac{2}{\pi}\right),\ \theta = \frac{k\pi}{16},\ k = 0,1,2,\ldots,96. \tag{4-6}$$

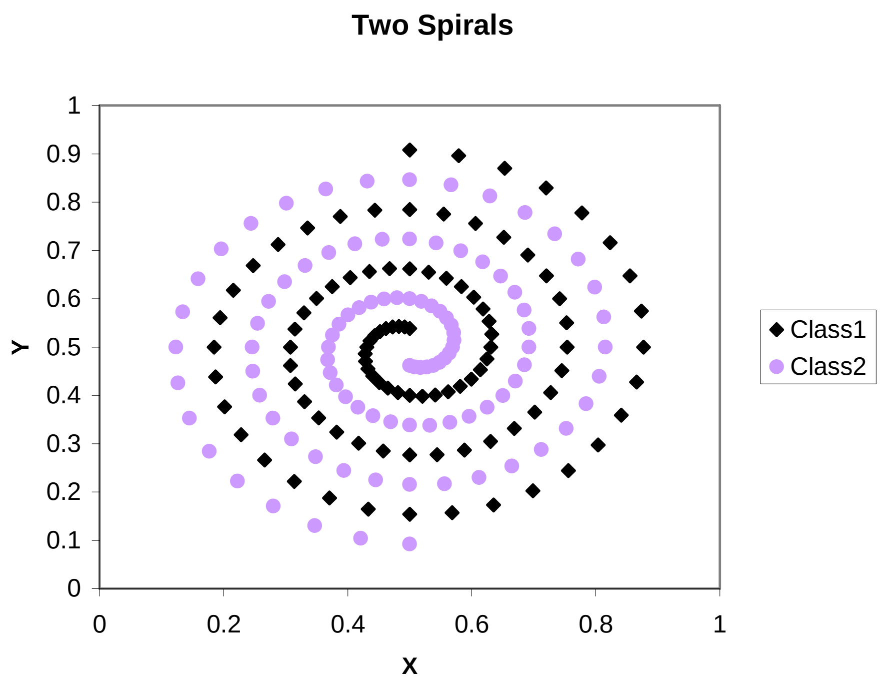


Figure 4-5: Two-Spiral dataset

The performance of the proposed HeRR system is compared with the Lang's 2-5-5-5-1 structure network (Lang and Witbrock 1988), Fuzzy ARTMAP (Carpenter, Grossberg et al. 1992), GenSoFNN and RSPOP. The experimental results are shown in Table 4-2.

Models	Training accuracy (%)	Testing accuracy (%)
Lang's 2-5-5-5-1 structure	100.0	92.8
Fuzzy ARTMAP	100.0	100.0
GenSoFNN	100.0	100.0
RSPOP	100.0	100.0
HeRR	100.0	100.0

Table 4-2: Experimental results on the Two-Spiral classification problem.

Lang and Witbrock (Lang and Witbrock 1988) proposed a special neural network with a 2-5-5-5-1 structure for this problem. It has 138 trainable weights, which are updated using vanilla backprogragation algorithm (Rumelhart, Hinton et al. 1986), with an average of 20000 training epochs. It yields the training accuracy of 100% and the testing accuracy of 92.8%.

Carpenter et al. (Carpenter, Grossberg et al. 1992) reports the results on the Two-Spiral dataset using the fuzzy ARTMAP system. It yields 100% training and testing accuracy by creating 194 ART categories for the 194 training samples. However, no fuzzy rules can be directly extracted from the fuzzy ARTMAP system.

Tung and Quek (Tung and Quek 2002) apply the GenSoFNN system to the Two-Spiral classification task. It is able to achieve 100% accuracy on both training set and testing set, using a total of 156 derived fuzzy rules. The RSPOP achieves the same accuracy. However, it needs 192 rules for classification.

The proposed HeRR system yields the 100% accuracy on the training and testing sets. It outperforms the systems mentioned above in the following aspects: (1) It does not need any training epochs, and the training data is fed into the system only in one pass. (2) A total of 74 fuzzy rules are generated, which is much less than that of GenSoFNN.

4.3.1.3 Iris classification

The Fisher's Iris dataset consists of 150 instances of Iris flowers belonging to three classes: namely, Sentosa (class 1), Virginica (class 2), and Versicolor (class 3). There are 50 instances for each of the three classes. Each instance consists of four attributes: namely, sepal length, sepal width, petal length, and petal width. The training set consists of 51 samples, with 17 samples from each class. The testing set has 99 samples, with 33 samples from each class. Three-fold cross-validation is used to evaluate the performance of the systems.

The proposed HeRR system is compared with the GenSoFNN (Tung and Quek 2002) , RSPOP (Ang and Quek 2005) and the FCMAC-BYY (Nguyen, Shi et al. 2006). The experimental results are shown in Table 4-3.

In Table 4-3, the classification accuracy on the three CV groups, as well as the overall accuracy is listed. The proposed HeRR system yields better accuracy on the CV1 and CV2 than that of GenSoFNN, RSPOP and FCMAC-BYY, where the accuracy on CV1 group is 100%. The accuracy on CV3 group of the four systems is the same. Thus, for the overall accuracy, the proposed HeRR system outperforms the other three systems.

Models	Classification accuracy on Iris dataset (%)				
	CV1	CV2	CV3	Mean	Std.
GenSoFNN	98.99	93.14	95.96	96.03	2.93
FCMAC-BYY	97.97	95.96	95.96	96.60	1.16
RSPOP	96.97	95.96	95.96	96.30	0.58
HeRR	100.0	98.99	95.96	98.32	2.10

Table 4-3: Experimental results on Iris dataset.

The parameter settings of the above three experiments are shown in Table A-5, in appendix A.

4.3.2 Experimental results on RS-HeRR for high-dimensional classification problems

The performance of the system is evaluated on five binary classification datasets: (1)

Pima Indian Diabetes dataset; (2) Urban water treatment plant monitoring dataset; (3) Sonar dataset; (4) Ovarian Cancer dataset; (5) Central Nervous System dataset.

4.3.2.1 Pima Indian Diabetes dataset

The Pima-Indian Diabetes dataset, downloaded from the UCI machine learning database repository (http://www.ics.uci.edu/~mlearn/MLRepository.html), consists of a total of 768 data instances, all of which are for the patients who are female of at least 21 years old Pima Indian heritage. There are 8 input features for each instance. They are listed in Table 4-4. It is a binary classification problem to classify whether the patient is tested as positive for diabetes.

To evaluate the performance of the RS-HeRR system, several other systems are employed to ensure a fair comparison. They are the Support Vector Machine (SVM) (Vapnik 1995), the C4.5 decision tree algorithm (Quinlan 1993), the Naïve Bayesian classifier (NB) (Friedman, Geiger et al. 1997), and the Rough-Set-based Pseudo Outer-Product fuzzy rule identification algorithm (RSPOP) (Ang and Quek 2005). The SVM, C4.5 and NB used in the experiment are implemented by the Weka software, which is a collection of machine learning algorithms for data mining tasks. It can be downloaded at http://www.cs.waikato.ac.nz/ml/weka/ .

10-fold cross-validation (CV) is involved to evaluate the performance of the system. All the 768 data points are divided approximately equally into 10 disjoint sets. In each CV group, one of the sets is used as the training set and the remaining are used as testing set. In other words, the ratio of the training and testing set is 1:9. The purpose of using the smaller subset of data as the training set is to test and compare the generalization ability of the proposed system and other benchmarked models.

Feature	**Description**
1	Number of times pregnant
2	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3	Diastolic blood pressure (mm Hg)
4	Triceps skin fold thickness (mm)
5	2-Hour serum insulin (mu U/ml)
6	Body mass index (weight in kg/(height in m)^2)
7	Diabetes pedigree function
8	Age (years)

Table 4-4: The description of the 8 features of the Pima Indian Diabetes dataset.

Models	Pima Indian Diabetes Dataset					
	Accuracy %		Sensitivity %		Specificity %	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	74.13	0.78	84.78	4.66	52.51	8.50
SVM	73.51	3.95	89.14	5.15	44.72	17.26
C4.5	73.03	4.62	81.88	9.48	56.88	13.89
NB	73.50	1.44	84.40	5.04	53.26	9.79
RSPOP	69.39	2.49	80.09	5.51	49.42	8.48

Table 4-5: Comparison on the Pima Indian Diabetes dataset.

Models	Pima Indian Diabetes dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Decrease %	Mean	Std.	Decrease %
HeRR	71.86	1.72	43.00	21.29		7.90	0.31	
RS-HeRR	74.13	0.78	17.90	19.79	58.37	3.10	1.91	60.76
HeRR+QuickReduct	73.35	1.49	32.90	26.42	23.49	3.80	2.04	51.90

Table 4-6: Comparison on the Pima Indian Diabetes dataset between HeRR, RS-HeRR, and HeRR with QuickReduct.

Besides the overall testing classification accuracy, another two measures, the sensitivity and the specificity, are used. The sensitivity and specificity are commonly used to evaluate the binary classifiers, especially in medical test. For example, sensitivity is the proportion of true positives of all diseased cases in the population. It can be seen as the probability that the test is positive given that the patient is sick. Specificity is the proportion of true negatives of all negative cases in the population. It can be considered as the probability that the test is negative given that the patient is not sick. To better analyze the performance of the systems in high-dimensional binary classification problems, especially for the two medical datasets, the Pima Indian Diabetes and the Ovarian Cancer, these two measures are employed in the comparisons. They are computed using Eqs. (4-3) and (4-4).

$$\text{Sensitivity}=\frac{\text{Number of positive samples correctly predicted}}{\text{Total number of positive samples}} \tag{4-3}$$

$$\text{Specificity}=\frac{\text{Number of negative samples correctly predicted}}{\text{Total number of negative samples}} \tag{4-4}$$

Table 4-5 shows the benchmark results of the RS-HeRR against the SVM, C4.5, NB, and RSPOP. The RS-HeRR system achieves the highest overall testing accuracy among other systems. Its standard deviation of the accuracy over the ten cross-validation groups is much lower than others, though its sensitivity is lower than that of SVM and its specificity is worse than that of C4.5 and NB. It indicates that, the generalization performance of the proposed RS-HeRR is better than that of SVM, C4.5, NB and RSPOP, when the number of training samples is much less than the number of testing samples.

To evaluate the performance of the proposed rough-set based attribute selection algorithm, another comparison is made where the HeRR system without rough set and the HeRR system plus the QuickReduct (Shen and Chouchoulas 2002) are used in the experiment. The simulation results are shown in Table 4-6. The overall testing accuracy, the number of generated rules and the number of selected features are compared in the table. The classification accuracy of the HeRR without rough set is the worst among these models, and the number of rules and features are much higher than others. It indicates that the redundancy in the rule set may degrade the testing accuracy of the HeRR system and the removal of redundant rules and features is necessary to maintain and boost the performance. The RS-HeRR system achieves a decrease of 58.37% for the number of rules and 60.76% for the number of features, compared with the HeRR system without any feature selection process. The decrease is larger than the HeRR plus QuickReduct. It indicates the rule set derived by RS-HeRR is more compact than that derived by HeRR plus QuickReduct. The performance of HeRR plus QuickReduct is closer to that of the RS-HeRR than others. That is because of the similarity of their rough-set based attribute selection method, where they all use the forward selection scheme. However, the proposed selection algorithm uses both the accuracy and the dependency degree to guide the selection process and the QuickReduct only employs the dependency degree. The experimental result shows the superior performance of the proposed rough-set based attribute selection algorithm.

One of the CV groups is used as an example to illustrate the operations of the proposed attribute selection algorithm. It is described as below:

After the Hebbian rule reduction algorithm (including rule initialization, rule ranking, fuzzy set merging and redundancy removing), a total of 17 fuzzy rules are derived. The classification accuracy on the training set and testing set are 80.52% and 71.35% respectively. At the beginning of the attribute selection algorithm, the attribute set A consists of all the attributes $x_1 - x_8$, and the selected attribute set R is $\varnothing$. For each attribute a in $A-R$, the training accuracy and the partial dependency using the rule set with the attribute set $R \cup \{a\}$ are computed (see Table 4-7).

Attributes $A-R$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Accuracy %	63.64	77.92	61.04	38.96	61.04	40.26	42.86	62.34
Dependency	0.00	0.67	0.33	0.33	0.00	0.50	0.20	0.20

Table 4-7: The training accuracy and partial dependency using $x_1 - x_8$

From Table 4-7, when the attribute x_2 is added into R, the training classification is the best among all the 8 attributes. Thus, R becomes $\{x_2\}$. As the training accuracy is worse than that using the whole attribute set, the selection process continues. In the second round of the loop, the training accuracy and partial dependency of the remaining 7 attributes are shown in Table 4-8. In the table, both the x_6 and x_8 achieves the best in training accuracy. However, the partial dependency of x_6 is larger than that of x_8. Thus the attribute x_6 is selected.

Attributes $A-R$	x_1	x_3	x_4	x_5	x_6	x_7	x_8
Accuracy %	77.92	77.92	76.62	76.62	79.22	77.92	79.22
Dependency	0.96	1.00	1.00	1.00	1.00	1.00	0.96

Table 4-8: The training accuracy and partial dependency using $x_1, x_3, x_4, x_5, x_6, x_7, x_8$.

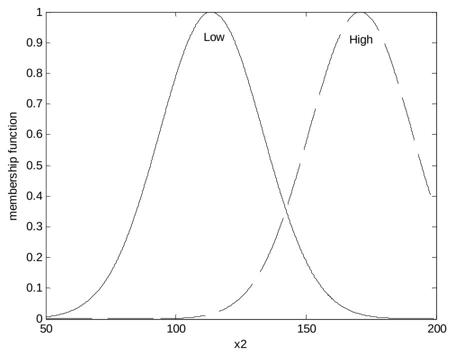
The algorithm goes to the third iteration, since the training accuracy is still lower than 80.52% which is the accuracy using the entire attribute set. In this iteration, the training accuracy and partial dependency are shown in Table 4-9. As the best accuracy is achieved by adding the x_7 into R, x_7 is selected. At this time, the accuracy is not lower than 80.52% and the partial dependency becomes 1. Thus the algorithm terminates.

Attributes $A-R$	x_1	x_3	x_4	x_5	x_7	x_8
Accuracy %	77.92	76.62	76.62	79.22	80.52	77.92
Dependency	1.00	1.00	1.00	1.00	1.00	1.00

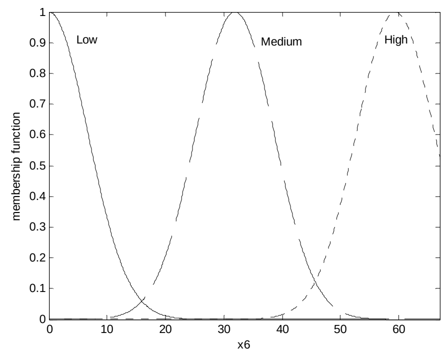
Table 4-9: The training accuracy and partial dependency using $x_1, x_3, x_4, x_5, x_7, x_8$.

At the end of the algorithm, the attribute set $\{x_2, x_6, x_7\}$ is produced. Using the 3 attributes, the rule set is simplified into 6 rules by deleting repeated rules. The training accuracy is still maintained at 80.52%. However, the testing accuracy is boosted to 75.40%, which is better than before, due to the selection of the attributes.

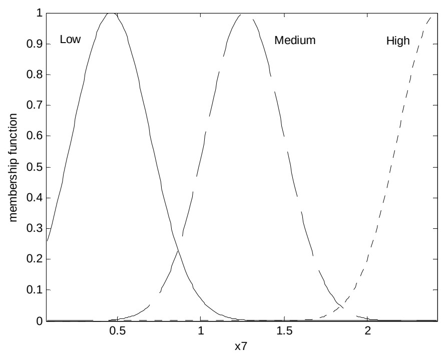
The interpretability of the proposed RS-HeRR system is evaluated in Figure 4-6 and Table 4-10. The Figure 4-6 shows the membership functions of the attributes x_2, x_6 and x_7. The x_2 is divided into 2 MFs (Low and High), and the x_6 and x_7 are divided into 3 MFs (Low, Medium and High). It is shown that the divided MFs are all clearly separated and have distinguished semantic meanings. The resultant rule set which consists of a total of 6 fuzzy rules is shown in Table 4-10. From the rule set, it can be seen that, when x_2 (Plasma glucose concentration) is low, x_6 (Body mass index) and x_7 (Diabetes pedigree function) are not high, the patient will be tested as positive for diabetes. When x_2 (Plasma glucose concentration) is high and x_6 (Body mass index) is not low, the patient will be tested as negative for diabetes. The derived rule set can be easily understood by humans and used to assist the decision making of the clinicians.



(a) Membership functions for x_2



(b) Membership functions for x_6



(c) Membership functions for x_7

Figure 4-6: Derived fuzzy membership functions of attributes x_2, x_6 and x_7.

Rule 1:	IF x_2 is *Low* and x_6 is *Medium* and x_7 is *Medium* THEN y is positive class
Rule 2:	IF x_2 is *Low* and x_6 is *Medium* and x_7 is *Low* THEN y is positive class
Rule 3:	IF x_2 is *Low* and x_6 is *Low* and x_7 is *Low* THEN y is positive class
Rule 4:	IF x_2 is *High* and x_6 is *Medium* and x_7 is *Low* THEN y is negative class
Rule 5:	IF x_2 is *High* and x_6 is *Medium* and x_7 is *Medium* THEN y is negative class
Rule 6:	IF x_2 is *High* and x_6 is *High* and x_7 is *High* THEN y is negative class

Table 4-10: The 6 derived fuzzy rules

4.3.2.2 Urban water treatment plant monitoring dataset

The urban water treatment dataset, downloaded from the UCI machine learning database repository, consists of a set of historical data measured from an urban waste water treatment plant over a period of 527 days. The objective is to classify the operational state of the plant in order to predict faults through the state variables of the plant at each of the stages in the treatment process. There are in total of 38 input features for each data point, organized into 5 aspects (Table 4-11).

Aspect	Number of features
Input to plant	9
Input to primary settler	6
Input to secondary settler	7
Output from plant	7
Plant performance	9

Table 4-11: The 5 aspects of the input features.

The status of the plant is represented by one of the 13 different categories. Some represent the normal operation of the plant and others point out various faults of the plant. They are listed in Table 4-12. As all the faults appear in a short period and are dealt with immediately, there are not enough training data points for these categories. To increase the number of data points for the faults, it is necessary to group fault cases together. In this experiment, all the data points are divided into to two major classes, 513 data for normal performance of the plant and the remaining 14 data for abnormal cases. In this dataset, there are missing values for some features. The missing values are replaced by the average value of the corresponding attribute in this experiment.

Similar to the previous experiment, 10-fold cross-validation is adopted. The whole

527 data points are divided approximately equally into 10 disjoint sets, where one of them is used as the training set and the other 9 sets are used for testing in each CV group.

Categories	Number of samples	Description
1	275	Normal situation
2	1	Secondary settler problem, type 1
3	1	Secondary settler problem, type 2
4	4	Secondary settler problem, type 3
5	116	Normal situation with performance over the mean
6	3	Solids overload, type 1
7	1	Secondary settler problem, type 4
8	1	Storm, type 1
9	69	Normal situation with low influent
10	1	Storm, type 2
11	53	Normal situation
12	1	Storm, type 3
13	1	Solids overload, type 2

Table 4-12: The 13 statuses of the water treatment plant.

The experimental results are shown in Table 4-13, benchmarked against SVM, C4.5 and RSPOP again. The RSPOP achieves a sensitivity of 1.0 with standard deviation of 0.0. It means that the RSPOP model correctly classifies all the positive samples (normal case) in all CV groups. However its performance on classifying negative samples (abnormal case) is bad, as its specificity is lower than others. In the experiment, the RS-HeRR achieves the best specificity. It makes the RS-HeRR system outperform other models in overall testing accuracy, although its sensitivity is slightly lower than that of SVM, C4.5 and RSPOP. In this dataset, the number of positive samples is much larger than that of the negative samples. The experimental results show the classification capability of the RS-HeRR in the dataset where the classes are unbalanced distributed.

Table 4-14 shows the comparison between the proposed rough set based attribute reduction algorithm and the QuickReduct (Shen and Chouchoulas 2002). Initial rule set is generated by the HeRR algorithm. The initial rule set without attribute reduction is also used for comparison. The testing accuracy achieved by RS-HeRR is the same as that of HeRR without RS, while the HeRR plus QuickReduct achieve lower accuracy. In the percent of decrease of the number of rules and features, compared with the HeRR without RS, the proposed RS-HeRR outperforms the HeRR with QuickReduct. It indicates that the RS-HeRR derives a more compact rule set while still maintaining the testing accuracy. On the other hand, just as the previous experiment where the training set is much smaller than the testing set, the RS-HeRR exhibits excellent generalization ability in pattern classification problems.

Models	Urban Water Treatment Dataset					
	Accuracy %		Sensitivity %		Specificity %	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	97.70	0.49	99.94	0.10	16.77	13.24
SVM	97.57	0.47	99.96	0.09	10.70	14.14
C4.5	97.47	0.45	99.98	0.69	6.06	12.86
NB	97.34	0.51	99.83	0.54	7.27	11.97
RSPOP	97.41	0.31	100.00	0.00	2.45	5.37

Table 4-13: Comparison on the Urban Water Treatment dataset.

Models	Urban Water Treatment dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Decrease %	Mean	Std.	Decrease %
HeRR	97.70	0.52	27.50	13.67		23.60	8.62	
RS-HeRR	97.70	0.49	3.50	3.87	87.27	1.10	0.32	95.34
HeRR+QuickReduct	97.68	0.49	4.80	7.27	82.45	1.20	0.42	94.92

Table 4-14: Comparison on the Urban Water Treatment dataset between HeRR, RS-HeRR, and HeRR with QuickReduct.

4.3.2.3 Sonar dataset

The Sonar dataset, downloaded from the UCI machine learning database repository, consists of a total 208 pattern samples with 60 input features. It is a binary classification problem where 111 samples of them are sonar returns from metal and 97 are sonar returns from rock. The dataset has been used in (Gorman and Sejnowski 1988; Kwak and Choi 2002). 3-fold cross-validation is used to evaluate the performance. The whole dataset is divided approximately equally into 3 disjointed sets. In each CV group, one set is used as the training set and the rest is used for testing.

Models	Sonar Dataset					
	Accuracy %		Sensitivity %		Specificity %	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	89.41	1.75	86.62	1.53	92.11	4.56
SVM	83.17	3.06	79.36	2.13	86.49	5.41
C4.5	78.82	4.79	67.93	9.91	88.29	4.13
NB	78.35	5.10	80.40	1.99	76.58	7.80
RSPOP	79.32	2.33	84.57	8.16	74.78	6.80

Table 4-15: Comparison on the Sonar dataset.

Models	Sonar Dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Decrease %	Mean	Std.	Decrease %
HeRR	90.86	2.23	135.33	1.15		60.00	0.00	
RS-HeRR	89.41	1.75	127.67	5.51	5.66	11.00	4.58	81.67
HeRR+QuickReduct	88.94	0.79	118.00	5.57	12.80	5.33	0.58	91.11

Table 4-16: Comparison on the Sonar dataset between HeRR, RS-HeRR, and HeRR with QuickReduct.

Table 4-15 shows the simulation results of RS-HeRR, compared with SVM, C4.5, NB, and RSPOP. The proposed RS-HeRR system yields an accuracy of 89.41±1.75%. The standard deviation of both the accuracy and sensitivity is lower than that of other models. In indicates that classification capability of the proposed RS-HeRR is better than the others in this dataset.

The comparison of the QuickReduct and the proposed rough-set based attribution algorithm is shown in Table 4-16. The HeRR without the RS attribute selection yields the highest testing accuracy, while the performance of the proposed RS-HeRR is in the second place. If there is no attribute reduction process, the HeRR will use all the 60 input features in the classification, and the derived fuzzy rules will be difficult to interpret. The RS-HeRR only involves 11±4.58 % input features, at a cost of about 1% decrease of accuracy. The results show that, a balance between accuracy and interpretability is achieved by the proposed RS-HeRR. The HeRR with QuickReduct is a competitive opponent to the RS-HeRR. Although its testing accuracy is slightly lower than that of RS-HeRR, the number of both derived rules and selected features is less than that of RS-HeRR.

4.3.2.4 Ovarian Cancer dataset

The ovarian cancer DNA microarray gene expression dataset (Schummer, Ng et al. 1999) consists of a total of 54 patterns where 30 examples are derived from ovarian tumors and 24 examples are normal. Each of these examples comprises 1536 features. The dataset has been used in (Li, Campbell et al. 2002; Tan, Quek et al. 2005). 3-fold cross-validation is used, similar with the previous experiment.

Models	Ovarian Cancer Dataset					
	Accuracy %		Sensitivity %		Specificity %	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	96.30	3.21	100.00	0.00	93.94	5.25
SVM	90.74	8.49	91.67	14.43	90.00	17.32
C4.5	90.74	3.21	91.67	7.22	90.00	10.00
NB	90.74	3.21	95.83	7.22	86.67	5.77

Table 4-17: Comparison on the Ovarian Cancer dataset.

Models	Ovarian Cancer Dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Decrease %	Mean	Std.	Decrease %
HeRR	70.37	6.41	36.00	0.00		1495	58.20	
RS-HeRR	96.30	3.21	12.33	5.77	65.75	2.67	0.58	99.82
HeRR+QuickReduct	96.30	3.21	17.67	9.29	50.92	3.33	1.53	99.78

Table 4-18: Comparison on the Ovarian Cancer dataset between HeRR, RS-HeRR, and HeRR with QuickReduct.

In Table 4-17, the RS-HeRR yields an accuracy of 96.30±3.21%. The SVM, C4.5 and NB yield the same average accuracy, 90.74%, where the standard deviation of the accuracy of C4.5 and NB is the same. In the Table 4-18, it shows that, the proposed RS-HeRR selects only 2.67±0.58 features out of the whole 1536 features for classification. However, the testing accuracy is enhanced from 70.37±6.41 to

96.30±3.21, due to the rough-set based attributes selection algorithm. The performance of the RS-HeRR and HeRR with QuickReduct in the terms of testing accuracy is the same. However, the RS-HeRR selects and derives less number of features and less number of rules than that of HeRR with QuickReduct. The proposed attribute selection algorithm outperforms the QuickReduct, as the proposed algorithm makes the rule set more compact.

4.3.2.5 Central Nervous System dataset

Embryonal tumors of the central nervous system (CNS) represent a heterogeneous group of tumors about which little is known biologically, and whose diagnosis, based on morphologic appearance alone, is controversial (Pomeroy, Tamayo et al. 2002). The Central Nervous System dataset, downloaded from the Kent Ridge Biomedical Data Set Repository (http://research.i2r.a-star.edu.sg/rp/), consists of the gene expression data of 60 patients. 21 of them are survivors who are alive after treatment, while the remaining 39 are failures who succumbed to their disease. There are 7129 input features for each instance. In this experiment, 5-fold cross-validation is used to evaluate the performance. The SVM, Naïve Bayesian classifier (NB) and C4.5 decision tree are used for comparison.

The experimental results are shown in Tables 4-19 and 4-20. From Table 4-19 we can see that, the proposed RS-HeRR system is much better than that of the commonly used classifiers, SVM, C4.5 and Naïve Bayes. It indicates the RS-HeRR is applicable to the problems where the input dimension is particular large and the number of instances is very small relative to the dimension. In Table 4-20 we can see that, the average number of the selected attributes is only 2.6, which results in a 99.96% decrease from the HeRR without attribute reduction. In the comparison of the proposed attribute reduction algorithm and QuickReduct, the former achieves a better accuracy than the latter, where the latter generates less number of rules. The average number of the selected attributes of RS-HeRR is the same as that of HeRR+QuickReduct.

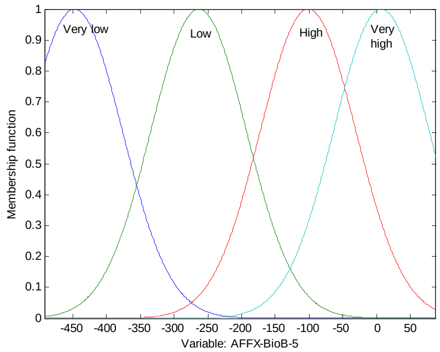
To illustrate the gained knowledge from learning, six derived rules from one of CV groups are shown in Figure 4-7 and Table 4-21. In this CV group, the genes AFFX-BioB-5, X66417 and U49020-cds2-s are selected for classification. There are 4, 5, and 5 membership functions derived for the three variables respectively. The derived membership functions are shown in Figure 4-7. We assign the fuzzy labels "Very low", "Low", "High", and "Very high" to the AFFX-BioB-5, and "Very low", "Low", "Medium", "High" and "Very high" to the other two variables. In Table 4-21, there are three rules for the survivors and three rules for the failures. The first rule says that if AFFX-BioB-5, X66417 and U49020-cds2-s are high, the patient will survive. In contrast to rule 4, if AFFX-BioB-5 and U49020-cds2-s are high but X66417 is low, there is a failure. It indicates that the X66417 classifies the patterns effectively when the other two are high. Similarly, the rule 2 in contrast to rule 5, the rule 3 in contrast to rule 6, indicate the other two variables are important under different conditions.

Models	Central Nervous System Dataset					
	Accuracy %		Sensitivity %		Specificity %	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	86.66	9.50	72.00	18.91	95.55	6.08
SVM	66.66	17.67	43.00	20.79	79.28	19.26
C4.5	56.67	14.91	39.00	34.17	67.50	18.96
NB	58.33	16.67	44.00	22.19	66.08	18.80

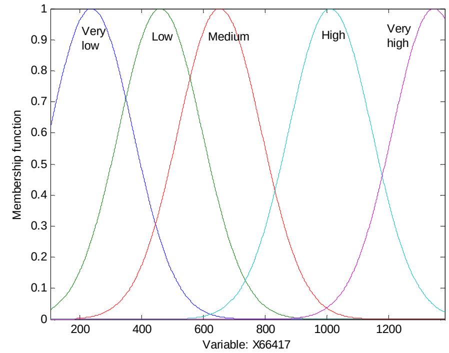
Table 4-19: Comparison on the Central Nervous System dataset.

Models	Central Nervous System Dataset							
	Accuracy %		No. of rules			No. of features		
	Mean	Std.	Mean	Std.	Decrease %	Mean	Std.	Decrease %
HeRR	49.99	21.24	48.00	0.00		7125.60	5.64	
RS-HeRR	86.66	9.50	35.60	8.35	27.08	2.60	0.54	99.96
HeRR+QuickReduct	79.99	4.56	32.00	5.74	33.33	2.60	0.54	99.96

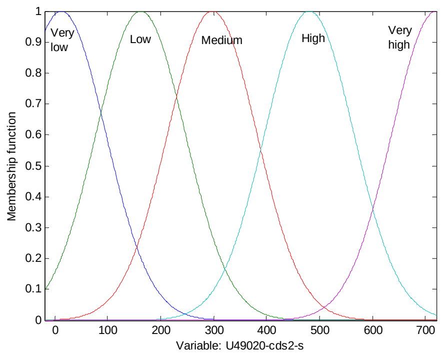
Table 4-20: Comparison on the Central Nervous System dataset HeRR, RS-HeRR, and HeRR with QuickReduct.



(a) Membership functions for AFFX-BioB-5



(b) Membership functions for X66417



(c) Membership functions for U49020-cds2-s

Figure 4-7: Derived fuzzy membership functions of attributes AFFX-BioB-5, X66417and U49020-cds2-s.

Rule 1	IF AFFX-BioB-5 is *High* and X66417 is *High* and U49020-cds2-s is *High* THEN the patient is survivor
Rule 2	IF AFFX-BioB-5 is *Very high* and X66417 is *High* and U49020-cds2-s is *Low* THEN the patient is survivor
Rule 3	IF AFFX-BioB-5 is *Low* and X66417 is *Low* and U49020-cds2-s is *High* THEN the patient is survivor
Rule 4	IF AFFX-BioB-5 is *High* and X66417 is *Low* and U49020-cds2-s is *High* THEN the patient is failure
Rule 5	IF AFFX-BioB-5 is *Very high* and X66417 is *Very low* and U49020-cds2-s is *Very low* THEN the patient is failure
Rule 6	IF AFFX-BioB-5 is *High* and X66417 is *Low* and U49020-cds2-s is *High* THEN the patient is failure

Table 4-21: Six derived fuzzy rules from one of CV groups.

The parameter settings of the above five experiments are shown in Tables A-6 to A-10, in appendix A.

4.4 Summary

This chapter presents the hybridization of HeRR neuro-fuzzy classifier and rough set theory, for generating fuzzy rules on the pattern classification problem. The rule set is initially generated using the HeRR (Liu, Quek et al. 2007). A post-processing step is used to further remove redundant attributes based on the rough set theory. In this step, a rough-set based attribute selection algorithm is proposed to reduce and simplify the rule set without affecting the classification accuracy. It uses both the system's performance and partial dependency as guides to determine suitable attributes subset.

The key strengths of the proposed methods are summarized as follows:

1) Effective and interpretable rule set. An effective and interpretable rule set is generated from the trained neuro-fuzzy system. The redundancy is removed from the rule set using rough-set based attribute selection algorithm, while the classification performance is maintained and boosted. Membership functions in each dimension are clearly separated and the size of the resultant rule set is small. It makes the rule set comprehensible to human beings.
2) Less user-defined parameters. The neuro-fuzzy system does not need to specify the number and positions of the membership functions in each dimension prior to the identification of rules. Only two free parameters are specified by the user in the whole system, i.e. the initial rule set generation threshold θ (in Section

3.1.1.) and the membership function merge threshold λ (in Section 3.1.2.).

The performance of the HeRR classifier is evaluated through three low-dimensional data sets: (1) the *Pat* synthetic pattern classification, (2) the Two-Spiral classification, and (3) the Iris classification. The performance of the RS-HeRR classifier is evaluated using five datasets: (1) the Pima Indian Diabetes dataset; (1) the Urban Water Treatment Plant Monitoring dataset; (3) the Sonar dataset; (4) the Ovarian Cancer dataset; (5) the Central Nervous System dataset. The experimental results show its superior performance when benchmarking against other established classifiers.

In the next chapter, two applications, the ICU Ventilation Modeling and the Back Failure Early Warning, will be introduced. Experimental studies are given to demonstrate the effectiveness of the proposed systems.

CHAPTER

5

5 Applications

"Success is the ability to go from
one failure to another with no loss of enthusiasm."
- Winston Churchill

The chapter discusses two applications, namely, (1) ICU Ventilation Modeling and (2) Back Failure Early Warning System. The former is a regression problem and is tackled by the Hebbian based Rule Reduction (HeRR) system, which is described in Chapter 3. The latter is a classification problem and is tackled by the Rough-Set based Hebbian Rule Reduction (RS-HeRR) system which is presented in Chapter 4. The experimental results show the superior performance of the HeRR and RS-HeRR systems on these real-world applications, against other well established systems.

5.1 ICU Ventilation Modeling

5.1.1 Background

Artificial ventilation plays a very important role in the treatment of patients in the Intensive Care Unit (ICU). In patients with impaired lung function, it provides a means to maintain the arterial oxygen and carbon dioxide levels. Plenty of research works on ICU ventilator weaning have been proposed in clinical research domain

(Bulter, Keenan et al. 1999; Weavind, Shaw et al. 2000; Huang and Lin 2006; Papadelis, Maglaveras et al. 2006). At hospitals, doctors in ICUs adjust the settings manually based on the patient status. The adjustment totally depends on clinical experience and expert knowledge. However, as new technologies are used, ventilators get more and more complex and it is not easy for clinicians with less experience to adjust them. Thus there is a need for the automation of the decision-making process and a need for a computer tool to help clinicians make a decision.

One of the first successful attempts to tackle with the ventilatory management task using artificial intelligence (AI) techniques is the ventilatory management expert system VM (Fagan, Kunz et al. 1979; Fagan 1980). The system is designed to interpret physiological state of the patients to infer appropriate ventilatory therapy. It uses AI knowledge representation techniques to the detection of possible artifacts and undesirable patient state.

Another expert system is the ESTER (Hernandez-Sande, Moret-Bonillo et al. 1989). It quantifies the delicacy of the patient's condition before the respiratory weaning process, using the APACHE-II criteria. The suggestions for the waning are offered from intermittent mandatory ventilation. It is developed using the GENIE, a knowledge engineering tool specially designed for building medical expert system. Static knowledge is organized in hierarchical frames, while the dynamic knowledge is embodied using sets of IF-THEN or IF-THEN-ELSE rules.

KUSIVAR (Rudowski, Frostell et al. 1989) is an expert system for mechanical ventilation of adult patients suffering from respiratory insufficiency. It provides guidance for respirator management during all phases of pulmonary disease using both qualitative rule-based knowledge and quantitative knowledge expressed in the form of mathematical models.

The VentEx system (Shahsavar, Gill et al. 1994) is a knowledge-based system consisting of a knowledge base including the domain knowledge represented by rule-based and object-oriented schemes and an inference engine including the mechanism for generating decision support. It is based on the KUSIVAR system, and integrates a domain knowledge specific tool, called KAVE (Shahsavar, Gill et al. 1991) , in the clinical environment using Nexpert Object.

Another ventilator management advisor, which combines both qualitative and quantitative computation, is the VentPlan (Rutledge, Thomsen et al. 1989; Rutledge, Thomsen et al. 1990; Rutledge, Thomsen et al. 1991). It employs a belief network to calculate the probability distributions of the shared physiologic model parameters from the qualitative and semi-quantitative inputs (Rutledge, Andersen et al. 1990; Polaschek, Rutledge et al. 1993). A mathematical model of cardiopulmonary physiology (Thomsen and Sheiner 1989) is implemented to predict the effects of alternative ventilator-control settings. The VentPlan ranks the proposed ventilator settings and their predicted effects using a plan evaluator, based on a multi-attribute-value model that specifies physician preferences for ventilator treatment. An expanded version of the VentPlan, the VentSim, has been developed

(Rutledge 1994). More detailed physiological model is employed in VentSim.

The NeoGanesh (Dojat, Pachet et al. 1997) is a knowledge-based system which controls the mechanical assistance provided to patients. It is a rule base system using a temporal reasoning model, and a closed-loop controller which has been tested in real medical situation.

Fuzzy system has been widely used in biomedical applications in recent years. It has been applied to control the ventilators in intensive care units (Seising, Schuh et al. 2003). In fuzzy modeling, the data can be interpreted in the term of linguistic terms, which are understandable to human user. The knowledge, in the form of fuzzy rules, can be extracted from the system, and provides the clinicians assistance in practical use. Expert's knowledge can also be incorporated into the system. In addition, fuzzy systems are more tolerable to the noises thus making them more robust. These advantages make the fuzzy system suited in the medical decision-making.

Fuzzy controller has been used to control the adjustment of inspired oxygen concentration (FiO2) for ventilated newborns using a set of fuzzy rules obtained from the expert knowledge of the neonatologists (Sun, Kohane et al. 1994), and the pressure support level for the patients with severe chronic obstructive pulmonary disease using a set of rules generated by the investigators (Nemoto, Hatzakis et al. 1999).

FuzzyKBWean is a knowledge-based fuzzy rule-base system for artificial ventilation

(Schuh, Hiesmayr et al. 1998; Schuh, Zelenka et al. 2000). It is an open-loop control system and designed to advise the change of the positive inspiratory pressure (PIP) level, the PEEP level, the inspiration time, the expiration time and the FiO2 level. The antecedents of the fuzzy rules consist of the linguistically expressed physiological parameters of the patients and actual ventilator settings, while their consequences contain the crisp values of the new settings of the ventilator.

The Fuzzy Advisor for Ventilator Management (FAVeM) is a fuzzy rule-base system, developed after the extensive literature survey and consultations with a clinical expert (Goode, Linkens et al. 1998). It is tested using a model of ventilation, namely the Simulation of Patients under Artificial Ventilation (SOPAVENT) (Goode 1993). They use the SOPAVENT as a physiological model (Kwok, Linkens et al. 2003), and employ the adaptive neuro-fuzzy inference system (ANFIS) (Jang 1993) to control the FiO2 (Kwok, Linkens et al. 2003). The results are validated by comparing with clinician's suggestions on the FiO2. Next, they create a new FiO2 advisory system, based on non-invasive estimation of the shunt using ANFIS (Kwok, Linkens et al. 2004). Finally, the, SIVA, a hybrid knowledge-and-model-based advisory system, is proposed (Kwok, Linkens et al. 2004). It consists of a top-level fuzzy rule-based module to give the qualitative component of the advice, and a lower-level model-based module to give the quantitative component of the advice. Closed-loop validation is performed in various medical scenarios.

5.1.2 ICU ventilator data

The data employed in the thesis is collected from the KK Women's and Children's Hospital of Singapore (Webpage: http://www.kkh.com.sg/). They represent a 20 day's records for a patient in the hospital under the BIPAP ventilation mode. The sampling time of the records is approximately one hour. All the time intervals may not be exactly one hour, because the measurements of the variables may take a little while. In the medical records, these amounts of time are usually several minutes. Compared with the one-hour time interval, it is small enough and will not affect much the accuracy and practicality of the results shown below. The records consist of measured patient-status variables and the setting variables by clinician. The patient-status variables include HR (Heart Rate), RR (Respiratory Rate), SaO2 (Oxygen Saturation), MAP (Mean Airway Pressure) and ETV (Expiratory Tidal Volume). The setting variables include FiO2 (Fraction of Inspired Oxygen), IE (Inspiratory / Expiratory Ratio), PEEP (Positive End Expiratory Pressure), PIP (Peak Inspiratory Pressure), RRset (set Respiratory Rate) and TVset (set Tidal Volume).

It is crucial to provide adequate oxygenation of the arterial blood for the maintenance of life. Thus, the arterial oxygen tension should be maintained of a proper level. The arterial oxygen tension is controlled by adjusting the FiO2. Among all of the variables in the medical records, the SaO2, FiO2, RR, PEEP, set or measured at current time step, are used to predict the value of FiO2 at next time step. To model the FiO2, the use of SaO2, FiO2 (old) and PEEP has been suggested (Kwok, Linkens et al. 2003).

The variable RR is advised by clinician.

5.1.3 Results and analysis

The objective of the experiment is to model the manual setting of FiO2 by clinician using a neuro-fuzzy hybrid system. A total of 408 data samples are divided into 3 cross-validation groups to evaluate the performance of the system. Within one cross-validation group, the first 60% data is used as training set and the following 40% data is used as testing set (see in Figure 5-1).

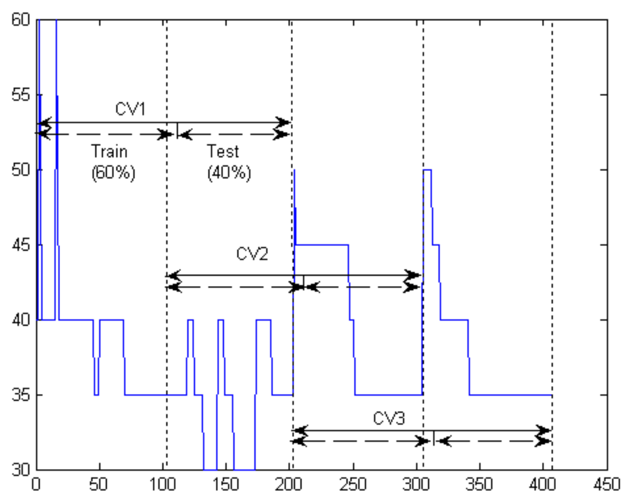


Figure 5-1: The FiO2 series and the division of 3 cross-validation groups.

In this experiment, the Root Mean Squared Error (RMSE) between the advised value of FiO2 by neuro-fuzzy hybrid systems, and the setting of the clinician is used to measure the performance. Some other neuro-fuzzy systems are employed for benchmark comparison. They are namely, EFUNN (Kasabov 2001), POPFNN (Ang, Quek et al. 2003), RSPOP (Ang and Quek 2005), DENFIS (Kasabov and Song 2002) and ANFIS (Jang 1993). The former 3 models are Mandani-type systems, while the latter 2 are TSK models. The Multi-Layer Perceptron (MLP) and Radial Basis Function network (RBF) are also used for comparison.

The experimental results are shown in Tables 5-1 and 5-2. Table 5-1 shows the comparison of RMSE between the above-mentioned neuro-fuzzy models. In both CV1 and CV2, the proposed HeRR system performs much better than other systems. In CV3, the ANFIS is slightly worse than the proposed system. The last column of Table 5-1 shows the average RMSE of the 3 cross-validation groups. The proposed system is the best among all of the neuro-fuzzy systems.

Table 5-2 shows the comparison of the number of the derived fuzzy rules. As the form of the TSK-style rule is different from the Mandani-style rule, these two kinds of rules are not comparable. Thus, the comparison is only made among the Mandani systems. The number of derived rules reflects the interpretability of the system. The more there are the rules, the more complex and the more uninterpretable the system is and vice versa. In Table 5-2, the proposed HeRR system produces the least number of rules in the CV1 and CV2. In CV3, only the RSPOP produce less number of rules than the HeRR. From the average number of rules in the last column, the proposed system produces the least among all these Mandani systems.

Figures 5-2 to 5-4 show the target and predicted values of FiO2 in testing set of the 3 cross-validation groups. In CV1, small deviations appear only when the amount of FiO2 becomes higher or lower abruptly. In CV2 and CV3, the output of the proposed system is nearly perfect-match to the target setting of FiO2.

The membership functions and some sample rules are extracted to show the interpretability of the proposed system. The membership functions for the 4 input variables and 1 output variable are shown in Figure 5-5. The overlap between the fuzzy sets is low for all the variables. Each fuzzy set has a clear semantic meaning.

Three samples fuzzy rules are shown in Table 5-3. These rules are the acquired knowledge from the data by the proposed system. They can be understood by human user and used to assist the clinicians.

The parameter settings of the results are shown in Table A-11, in appendix A.

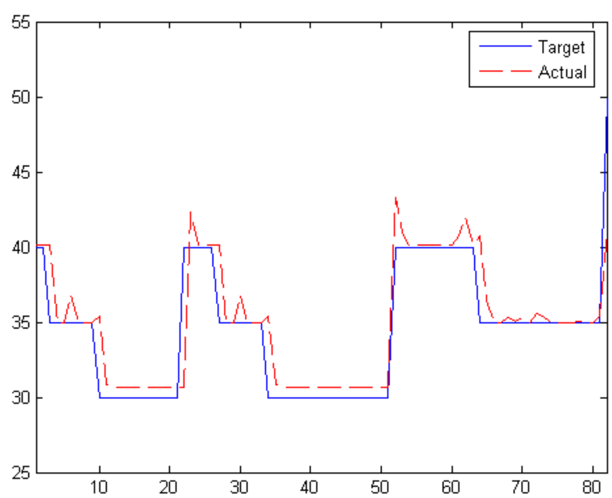


Figure 5-2: Target and predicted values of FiO2 in CV1

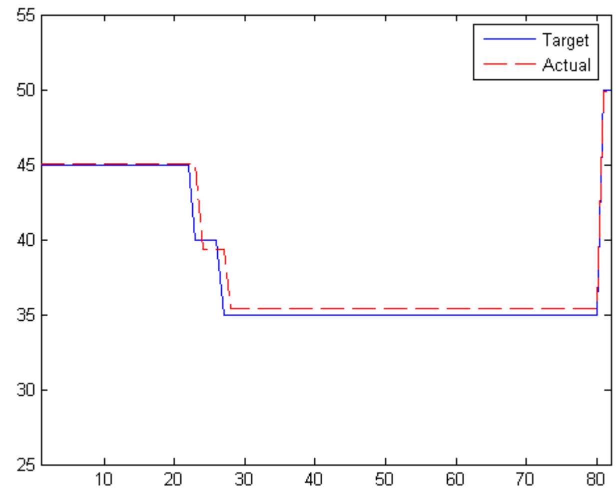


Figure 5-3: Target and predicted values of FiO2 in CV2

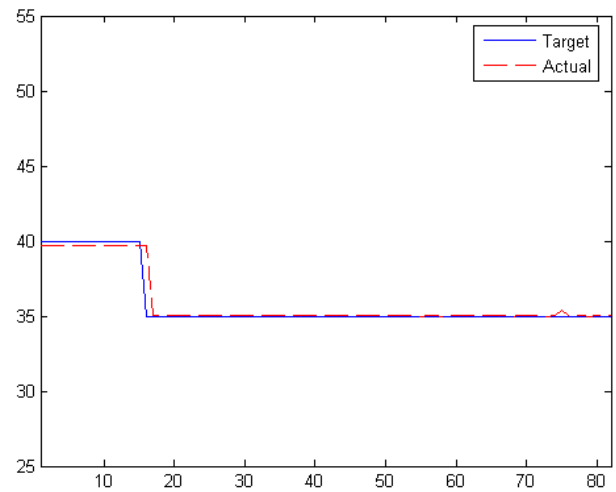


Figure 5-4: Target and predicted values of FiO2 in CV3

Models	RMSE				
	CV1	CV2	CV3	Mean	Std.
HeRR	2.085	0.753	0.536	1.125	0.839
POPFNN	13.776	6.031	2.375	7.394	5.821
RSPOP	13.841	6.031	2.166	7.346	5.948
EFuNN	3.417	2.908	1.219	2.515	1.151
DENFIS	3.045	2.243	1.045	2.111	1.007
ANFIS	2.409	1.863	0.560	1.611	0.950
MLP	2.940	2.822	0.654	2.139	1.238
RBF	6.416	1.768	2.220	3.468	2.563

Table 5-1: Testing accuracy on the three CV groups.

Models	No. of rules				
	CV1	CV2	CV3	Mean	Std.
HeRR	12	3	26	13.67	11.59
POPFNN	48	28	53	43.00	13.23
RSPOP	29	9	11	16.33	11.02
EFuNN	32	57	109	66.00	39.28

Table 5-2: The number of derived fuzzy rules (Mamdani type).

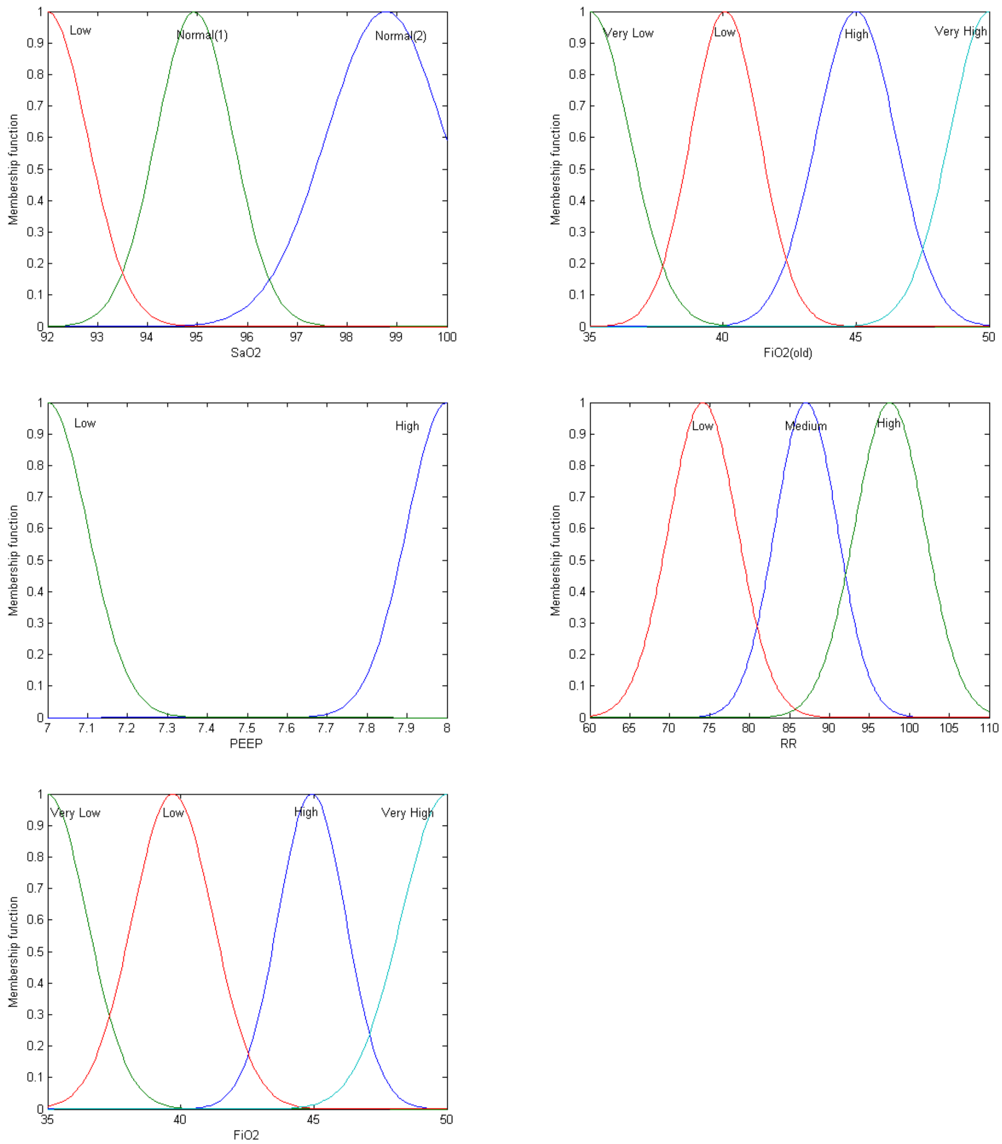


Figure 5-5: Fuzzy membership functions on SaO2, FiO2 (old), PEEP, RR and FiO2

Rule 1	IF	SaO2 is *Normal(2)* and FiO2 is *Medium* and PEEP is *High* and RR is *Medium*,
	THEN	the new FiO2 is *High*.
Rule 2	IF	SaO2 is *Low* and FiO2 is *High* and PEEP is *Medium* and RR is *High*,
	THEN	the new FiO2 is *High*
Rule 3	IF	SaO2 is *Normal(1)* and FiO2 is *High* and PEEP is *Medium* and RR is *Low*,
	THEN	the new FiO2 is *Medium*.

Table 5-3: Three sample derived fuzzy rules.

5.2 Bank Failure Early Warning System

5.2.1 Background

The collapse and failure of a bank could have devastating consequences to the entire banking system and an adverse repercussion effect on other banks and financial institutions. Some of the negative impacts are the massive bail out cost for a failing bank and the negative sentiments and loss of confidence developed by investors and depositors. Hence, bank failure prediction is an important issue for regulators of the banking industries.

Various traditional models based on statistical methodologies have been employed to study the problem, such as multivariate discriminant analysis (MDA) (Sinkey 1975), logit models (Martin 1977) and Cox's proportional hazards models (the Cox's model) (Lane, Looney et al. 1986; Cole and Gunther 1995). Neuro-fuzzy models are also employed for predicting bank failures. They are the GenSo-EWS (Tung, Quek et al. 2004), FCMAC-EWS (Ng, Quek et al. 2008) and FCMAC-BYY (Nguyen, Shi et al. 2006).

Category	Covariate	Expected impact on failure
Capital Adequacy	**CAPADE** average total equity capital (3210) / average total assets (2170)	Higher is the ratio, greater is the capacity to absorb losses, and smaller is the probability to failure.
Asset (loan) Quality	**OLAQLY** average (accumulated) loan loss allowance (3123) / average total loans & leases, gross (1400)	Smaller is the ratio, better is the loan quality, and smaller is the probability to failure.
	PROBLO (average (accumulated) loans 90 + (days late (1407)) / average total loans & leases, gross (1400)	Higher is the ratio, poorer is the loan quality, and higher is the probability to failure.
	PLAQLY (annual) loan loss provision (4230) / average total loans & leases, gross (1400)	Higher is the ratio, poorer is the loan quality expected to be, higher is the probability to failure.
Management	**NIEOIN** non interest expense (4093) / operating income (4000)	Higher is the ratio less operationally efficient profitable is the bank, higher is the probability to failure.
Earning	**NINMAR** (total interest income (4107) – interest expense (4073)) / average total asset (2170)	Higher is the return on equity before tax, smaller is the probability to failure.
	ROE (net income (after tax) (4340) + applicable income taxes (4302)) / average total equity capital (3210)	Higher is the return on equity before tax, longer is the time to regulatory closure or financial distress.
Liquidity	**LIQUID** (average cash (0010) + average federal funds sold (1350)) / (average total deposit (2200) + average fed funds purchased (2800) + average banks' liability on acceptance (2930) + average other liabilities (2930))	Higher liquidity indicates inefficient utilization of resources; it can also reflect an expectation of unfavorable events (runs on deposits for example). Overall, higher liquidity suggests a higher probability of failure.
Miscellaneous	**GROWLA** (total loans & leases, gross $(1400)_t$ - total loans & leases, gross $(1400)_{t-1}$) / total loans & leases, gross $(1400)_{t-1}$	With appropriate credit control and adequate loan loss provisions, a bank with higher loan growth rate would have better profitability and smaller probability to failure.

Table 5-4: Nine variables for bank failure prediction.

The financial variables (covariates) used in the bank failure prediction application are extracted from the Call Report, which are downloaded from the website of Federal Reserve Bank, Chicago (http://www.chicagofed.org). The expected impacts of the variables on bank failures are explained in (Cheng 2002). Normality plots of these variables indicate that the variables are not normally distributed. The statistical significance of the variables is investigated by *best score selection, stepwise selection and purposeful selection*. Based on the findings of these selection procedures and an analysis of the correlations between the variables, only the nine variables listed in Table 5-4 are incorporated for bank failure classification and prediction.

5.2.2 Experiment setup

The financial variables (covariates) used in the experiment is identical to those used in the Cox's Proportional Hazards Model (Lane, Looney et al. 1986; Cole and Gunther 1995). The observation period of the survived banks consists of 21 years from January 1980 to December 2000. There are 702 failed banks and 2933 survived banks over the observation period, leading to a total of 3635 banks. Based on statistical investigation, nine covariates are selected according to their significance and correlation (Table 5-4).

Three different scenarios of experiments are conducted:

1) Bank failure prediction based on the last available financial record
2) Bank failure prediction using financial records one year prior to the last one

3) Bank failure prediction using financial records two years prior to the last one

For the failed banks, the last available financial statement would be those prior to failure. The records for the surviving banks are those of the year 2000 (last year of the observation period). For the three sets of experiments, records on different years are extracted from the original data set. Those with missing data are filtered out. The number of banks records, survived banks and failed banks in each scenario are listed in Table 5-5. For each experiment, the data set is split into one training set and one testing set. The training set consists of 20% of the data set and the testing set contains the remaining 80% of the data set. Five cross-validation groups are used to evaluate the performance of the systems, denoted CV1 to CV5, respectively. The original data set is initially split into two pools: failed and survived banks. The number of samples of failed and survived banks in the training set is not balanced. For each cross-validation group, the training set is randomly selected from two pools so that the number of survived and failed bank is equal. It is called a "balance" training scenario. The training sets of the five groups are mutually exclusive. As a result of the "balance", the size of each training set is shortened to about 7% of the whole data set.

	Total	Survived	Failed
Last available record	3103	2555 (82.34%)	548 (17.66%)
One year prior	3046	2572 (84.44%)	474 (15.56%)
Two years prior	2943	2585 (87.84%)	358 (12.16%)

Table 5-5: Number of available records in each scenario.

5.2.3 Results and analysis

One output is used to differentiate between failed and survived banks. Failed banks are denoted with output “0” while survived banks are identified by output “1”. The proposed RS-HeRR system is subsequently used to model the inherent relationships between the financial covariates and their impact on the financial solvency of the respective banks. The RS-HeRR system is trained using the training set and the modeling capability of the trained network is subsequently evaluated using the testing set. The simulation is repeated for all the five cross-validation groups.

The performance of the RS-HeRR system is compared with the published results on the data set using the GenSo-EWS (Tung, Quek et al. 2004), FCMAC-EWS (Ng, Quek et al. 2008) and FCMAC-BYY (Nguyen, Shi et al. 2006). Classical approaches on classification problems, such as SVM, Naïve Bayes (NB) and C4.5, are also employed for comparison.

Tables 5-6 to 5-8 show the performance of the RS-HeRR, compared with GenSo-EWS, FCMAC-EWS, FCMAC-BYY, SVM, NB, and C4.5. From the tables, it can be seen that, the classification accuracy degrades with respect to the prediction period. The longer the prediction period, the less accurate the classification on the testing set. The proposed RS-HeRR fuzzy classifier outperforms other systems in all the CV groups of the three experiments. The RS-HeRR yields the mean accuracy of 97.68% in the first scenario, 94.52% in the second scenario, and 93.47% in the last scenario. The most competitive results are achieved by SVM. The performance of

RS-HeRR is slightly better than that of SVM. Although the improvement is not quite significant, the RS-HeRR has generated interpretable rules (Table 5-14) from data, which is an advantage over SVM.

Tables 5-9 to 5-11 show the performance of the proposed attribute reduction algorithm. It is conducted by using the same rule generation method (the HeRR) with different or without attribute reduction algorithm. The first row of the three tables is the performance of the RS-HeRR, including the classification accuracy, the number of rules and the number of employed attributes. The second and the third rows are for the HeRR with QuickReduct algorithm, and the HeRR without attribute reduction, respectively. The mean value and its standard deviation of all the five cross-validation groups are presented for each experimental scenario. From these tables, we can see that, the proposed attribute algorithm does reduce redundancy from the derived fuzzy rule set by HeRR system. The number of the resultant rule set and the number of employed attributes are much smaller than that without attribute reduction. Furthermore, the accuracy of all the three scenarios are improved, which indicates that the reduction of redundancy has surely enhanced the quality of the rule set. Compared with the QuickReduct algorithm, the proposed rough-set based attribute algorithm reduces more redundant attributes and rules, and yields better classification accuracy than that of the QuickReduct in all the three scenarios.

	Scenario 1: Last record classification rate (%)						
	CV1	CV2	CV3	CV4	CV5	Mean	Std.
RS-HeRR	97.74	97.70	97.58	97.46	97.90	97.68	0.17
GenSo-EWS	92.92	90.62	94.37	86.04	95.21	91.83	3.68
FCMAC-EWS	94.18	96.47	96.16	94.44	95.12	95.27	1.02
FCMAC-BYY	93.95	94.96	94.48	94.64	94.60	94.53	0.37
SVM	97.38	97.86	97.06	97.38	97.58	97.45	0.29
NB	96.29	96.53	96.41	96.98	97.30	96.70	0.42
C4.5	92.42	94.60	94.36	95.57	96.17	94.62	1.43

Table 5-6: The classification accuracy on the experimental scenario of last record.

	Scenario 2: One year prior classification rate (%)						
	CV1	CV2	CV3	CV4	CV5	Mean	Std.
RS-HeRR	94.41	94.59	94.54	94.54	94.50	94.52	0.07
GenSo-EWS	82.26	81.13	87.55	76.60	86.79	82.87	4.47
FCMAC-EWS	89.40	90.32	90.29	90.38	90.64	90.21	0.47
FCMAC-BYY	93.23	92.77	92.64	93.53	92.77	92.99	0.38
SVM	94.12	94.71	91.88	95.24	92.29	93.65	1.49
NB	90.99	93.85	92.16	93.07	91.88	92.39	1.10
C4.5	89.10	90.89	88.35	92.94	93.07	90.87	2.16

Table 5-7: The classification accuracy on the experimental scenario of one year prior.

	Scenario 3: Two years prior classification rate (%)						
	CV1	CV2	CV3	CV4	CV5	Mean	Std.
RS-HeRR	93.84	93.38	93.93	93.33	92.87	93.47	0.43
GenSo-EWS	68.29	68.29	81.95	74.15	68.29	72.19	10.25
FCMAC-EWS	84.60	84.22	83.30	86.66	80.00	83.76	2.43
FCMAC-BYY	91.73	91.81	91.19	92.06	92.14	91.79	0.37
SVM	93.32	93.37	93.71	91.93	92.82	93.03	0.69
NB	90.86	90.53	91.89	90.23	89.98	90.70	0.74
C4.5	89.96	86.92	91.04	89.94	88.07	89.18	1.66

Table 5-8: The classification accuracy on the experimental scenario of two year prior.

	Scenario 1 (%)		Scenario 2 (%)		Scenario 3 (%)	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	97.68	0.17	94.52	0.07	93.47	0.43
HeRR+QuickReduct	97.20	0.31	93.00	0.69	92.58	0.48
HeRR	96.46	0.37	92.41	1.16	92.26	0.34

Table 5-9: Comparison of the classification accuracy.

	Scenario 1		Scenario 2		Scenario 3	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	38.20	13.97	22.80	19.87	27.60	23.08
HeRR+QuickReduct	43.20	29.32	89.00	62.94	27.80	16.93
HeRR	115.80	54.66	104.6	39.69	51.40	45.46

Table 5-10: Comparison of the number of rules.

	Scenario 1		Scenario 2		Scenario 3	
	Mean	Std.	Mean	Std.	Mean	Std.
RS-HeRR	3.00	1.41	2.60	1.52	3.60	1.82
HeRR+QuickReduct	3.00	1.41	3.80	1.48	4.20	1.48
HeRR	8.20	0.45	9.00	0.00	7.80	0.84

Table 5-11: Comparison of the number of attributes.

The attribute sets selected by the proposed RS-HeRR system in these CV groups are shown in Table 5-12. The number 1-9 denotes the input attributes of the data set. That is, 1-CAPADE, 2-OLAQLY, 3-PROBLO, 4-PLAQLY, 5-NIEOIN, 6-NINMAR, 7-ROE, 8-LIQUID, and 9-GROWLA. The semantic meaning of the attributes is listed in Table 5-4. The frequency of the attributes used in all the CV groups is shown in Table 5-13. From the table we can see that, the attributes 1-CAPADE, 4-PLAQLY and 7-ROE are most frequently used in the experiments, where 7-ROE has the largest frequency. It indicates that, the attributes 1-CAPADE, 4-PLAQLY and 7-ROE have more impact on the bank failure than other attributes and should be paid more attention.

	Selected attribute set				
	CV1	CV2	CV3	CV4	CV5
Last record	(1,4,7,9)	(1,4)	(1,4,7,9)	(7)	(1,4,7,9)
1 year prior	(1,3,4,6)	(1,4,9)	(7)	(7)	(1,3,4,7)
2 year prior	(2,7)	(7,9)	(1,4,7,8,9)	(1,7)	(1,2,3,4,7,9)

Table 5-12: Selected attribute set by the RS-HeRR for all the CV groups.

Attribute	1	2	3	4	5	6	7	8	9
Frequency	10	2	3	9	0	1	12	1	7

Table 5-13: Frequency of the 9 attributes in all the CV groups.

The CV4 group in the third experiment scenario is used as an example to show the interpretability of the RS-HeRR system. In this CV group, two attributes are selected for predicting the failed banks, 1-CAPAD and 7-ROE. The membership functions (MF) of the two attributes are shown in Figure 5-6. There are 3 MFs for 1-CAPAD and 3 MFs for 7-ROE, with the semantic meanings *Low*, *Medium* and *High*. The derived rules are listed in Table 5-14. From the seven rules, we can see that, the first two rules are used to classify the failed banks, while the others are for the survived banks. It can be seen that, if the ROE is *High* and the CAPADE is not *Low*, then the bank will fail, otherwise the bank will survive. The derived rule set has shown a simple relationship between the attributes and the bank failure.

The parameter settings of the results are shown in Tables A-12 to 14, in appendix A.

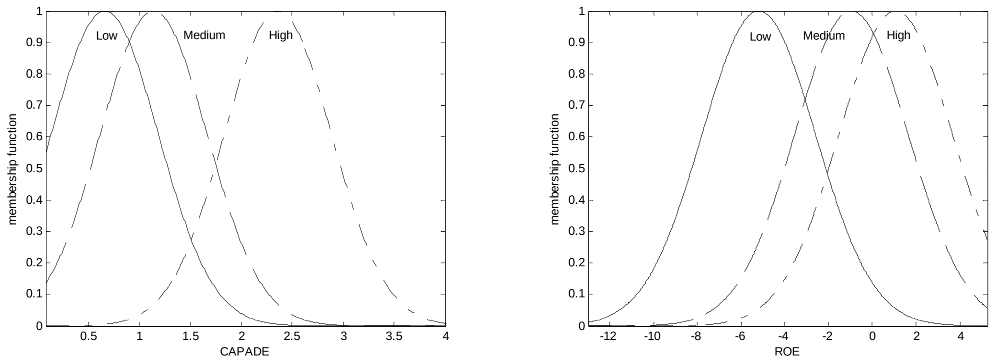


Figure 5-6: Membership functions of variables CAPADE and ROE

Rule	CAPADE	ROE	Class
1	*Medium*	*High*	*Failed*
2	*High*	*High*	*Failed*
3	*Low*	*Low*	*Survived*
4	*Low*	*Medium*	*Survived*
5	*Medium*	*Low*	*Survived*
6	*Medium*	*Medium*	*Survived*
7	*High*	*Medium*	*Survived*

Table 5-14: The derived rule set.

5.3 Summary

In Section 5.1, the setting of FiO2 in artificial ventilation is modeled by the Hebbian based rule reduction (HeRR) system. The proposed system is tested on the real ventilation data. The overall modeling error is lower than the other comparable neuro-fuzzy system, while the extracted fuzzy rules can be understood by human user and used to assist clinicians.

Section 5.2 has discussed the application of RS-HeRR to the bank failure prediction. Many statistical models such as the Cox's model (Lane, Looney et al. 1986; Cole and

Gunther 1995) have been applied to the study of bank failure. However, these models have not attempted to identify the possible traits of financial distress that leads to bank failure. In these models, it is difficult to explicitly specify what constitutes a financial distress and the intrinsic relationship between financial distress and a failed bank. In the section, the Rough-Set based Hebbian Rule Reduction neuro-fuzzy system is used to predict the bank failure. The selected attribute sets are analyzed to indicate what factor has more impact the bank failure. Fuzzy rules are derived to show the relationship of the attributes and the bank failure. The experimental results show superior performance of the proposed RS-HeRR system, benchmarked against other published bank failure early warning systems.

In the next chapter, the conclusion of the thesis will be presented, as well as the future work on this research topic.

CHAPTER

6

6 Conclusion and future work

"Don't judge each day by the
harvest you reap, but by the seeds you plant!"
- Robert Louis Stevenson

This Chapter presents the conclusion of the research work (see Section 6.1) and the future work (see Section 6.2).

6.1 Conclusion

In this chapter, the achievements done in this research work is summarized. The overview of the research work is shown in Figure 6-1.

As described in Figure 6-1, the objective of the thesis is to construct a generic linguistic neuro-fuzzy system that is able to generate interpretable fuzzy rules while maintaining acceptable modeling accuracy. To fulfill the objective, three issues are investigated in the thesis, based on the drawbacks of the existing neuro-fuzzy systems. They are namely, (1) Defining the MFs; (2) Knowledge reduction, including the reduction of inconsistent rules and redundant attributes; and (3) Trade-off between interpretability and accuracy.

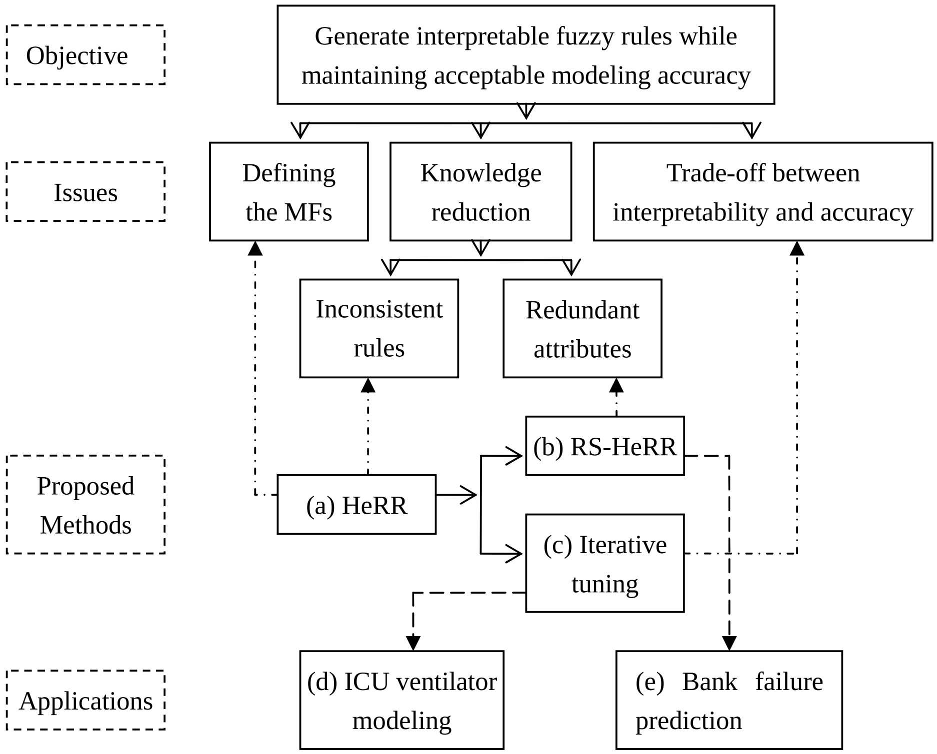


Figure 6-1: The research work discussed in the thesis.

To handles these issues, the proposed methods are concluded as follows:

a) A novel Hebbian based Rule Reduction (HeRR) algorithm is proposed to derive interpretable fuzzy membership functions and fuzzy rules. Existing methods to define the fuzzy membership functions, such as grid-type partitioning method (Cordon and Herrera 2000) (Kasabov 2001) and clustering algorithms (Baraldi and Blonda 1999; Baraldi and Blonda 1999; Xu and Wunsch II 2005), usually require prior knowledge on the MFs, such as the number of MFs. As the number of attributes increases, it is difficult to manually set them on each dimension and, the extensive searching on them will suffer from the computational inefficiency. The proposed HeRR algorithm tackles with the problem through the merger of similar MFs. Fuzzy rules with excessive fuzzy membership functions are initially generated from the training data. These MFs are merged through the proposed merging process. As a result, the input space is partitioned into several well-separated parts for each dimension. Each partition is assigned a semantic fuzzy label, with a Gaussian fuzzy membership function. The fuzzy rules are derived by linking the MFs together. The proposed HeRR algorithm provides an approach for an effective automatic generation of MFs without the use of apriori knowledge. Furthermore, it resolves the inconsistency in rule set using the Hebbian importance defined for each rule. The rule with higher Hebbian importance is retained when inconsistency occurs (Chapter 3).

b) An iterative tuning and reduction process is proposed to balance the accuracy and interpretability of the derived fuzzy rules for the regression problem. Interpretability and accuracy are two important issues in neuro-fuzzy modeling. However, these two requirements are usually hard to be satisfied concurrently. Trade-off between interpretability and accuracy is desirable, which is ignored by many existing neuro-fuzzy systems (see Section 2.3.3). The research work in the thesis deals with the problem through an iterative tuning and reduction process. It consists of two phases, tuning and reduction. In the tuning phase, the membership functions are tuned through the Least Mean Square (LMS) algorithm; in the reduction phase, the fuzzy sets are reduced through the proposed HeRR algorithm. Its performance on a set of experiments shows that it is able to generate interpretable fuzzy rules while maintaining acceptable modeling accuracy

(Chapter 3).

c) A novel Rough-Set based Hebbian Rule Reduction (RS-HeRR) system, which is a hybrid of HeRR and rough set, is proposed for the pattern classification problems. Many other neuro-fuzzy systems, reviewed in Section 2.3.3, mainly focus on the knowledge acquisition and construction with lesser attention to knowledge reduction. Redundant attributes do not only undermine the interpretability of the system but also affect the modeling accuracy. A rough-set based attributes reduction algorithm is proposed to remove redundant attributes. It starts with an empty set of attributes and selects attributes through the guild of the classification accuracy and the partial dependence between attributes. By eliminating the redundant attributes and rules, a more compact and interpretable rule set is derived while the accuracy does not decrease (Chapter 4).

d) An ICU artificial ventilator modeling problem is handled through the proposed HeRR neuro-fuzzy system. Interpretable fuzzy rules are derived, and the modeling accuracy is higher than other benchmarked systems (Chapter 5).

e) A bank failure prediction problem is tackled using the proposed RS-HeRR classifier. It achieves superior classification performance. The factors that have more impact on the bank failure and the relationship between the input attributes and the failed bank are also discussed (Chapter 5).

The proposed HeRR algorithm with iterative tuning and reduction is a generic approach to regression problems. It is able to approximate a number of functions and has been successfully applied to real-world applications. Besides the modeling accuracy, it is capable of generating compact and interpretable fuzzy rule set during training. It provides the transparency to the users. However, as a limitation, the current version of the system still runs in an off-line mode. The generalization to on-line learning needs further investigation. To generalize the HeRR to classification problems, the RS-HeRR is proposed. In RS-HeRR, redundant attributes are reduced through rough set theory to form a compact rule set without affecting the accuracy. It is suitable for both low-dimension and high-dimension binary classification problems (Chapter 4, Section 4.3). For multi-class classification problems, further research on the improvement of RS-HeRR is still needed. The RS-HeRR is able to achieve comparable and better results on these problems, compared to other neuro-fuzzy systems, such as RSPOP, GenSoFNN and EFuNN, and classical statistical classifiers, such as SVM, Naïve Bayes and C4.5 decision tree. Particular in some of the results, the performance of RS-HeRR is quite near to that of SVM. The merit of RS-HeRR over SVM is that it is able to derive a set of interpretable rules to explicitly represent the gained knowledge from learning.

As the major conclusion, the presented research works have successfully fulfilled the objective: Generate interpretable fuzzy rules while maintaining acceptable modeling accuracy, as well as the research issues described above. It has overcome the drawbacks of the existing neuro-fuzzy systems, described in Chapter 1, Section 1.1.4.

As the major contribution in the thesis, the proposed Hebbian based Rule Reduction algorithm with the iterative tuning and reduction process, described in Chapter 3, has been published in the *Neural Computation* journal (Liu, Quek et al. 2007). The result of the ICU ventilation modeling problem has been published in the *International Joint Conference on Neural Networks* (Liu, Quek et al. 2006).

6.2 Future work

The research of neuro-fuzzy modeling is grouped as two branches: the linguistic neuro-fuzzy modeling, which is implemented through the Mamdani-type fuzzy rules, and precise neuro-fuzzy modeling, which is realized through the TSK-type fuzzy rules. The linguistic modeling has been studied in the thesis. In the future work, the research will be extended to the TSK-type neuro-fuzzy systems. The form of the condition part of the TSK-type fuzzy rules is the same as that of Mamdani-type rules. Hence the proposed HeRR can be directly used to formulate the fuzzy membership functions in each input dimension. On the other hand, the consequent part of the TSK-type rules is a linear combination of the input variables, instead of a fuzzy set. The improvement of the interpretability on the consequent part of TSK-type rules has become a challenging research topic (Casillas, Cordon et al. 2003). The extension of the current work or the proposal of new algorithms to tackle with this problem is a direction in further research.

In this thesis, the problem of attribute reduction has been investigated. The proposed method is used to simplify the fuzzy rule set based on the rough set theory. In recent research on attribute selection, the Mutual Information (MI) has been considered a good indicator of relevance between variables that is robust to noise and data transformation. A series of research efforts have been made in this area (Kwak and Choi 2002; Chow and Huang 2005; Peng, Long et al. 2005). The development of new attribute selection algorithm based on the hybrid of rough set and MI is another interesting research topic.

In the future work, new applications will be considered. The financial forecasting is one of fascinating research areas. The proposed method in the thesis can be used either to predict the daily stock price or to construct trading rules from historical financial data. In recent years, AI approaches have been used to develop intelligent trading system (Ang and Quek 2006). Among these approaches, reinforcement learning is a promising learning scheme that is able to automatically learn the trading strategy to maximize accumulated profits in the long run (Moody, Wu et al. 1998; Moody and Saffell 2001; Dempster and Leemans 2006; Jangmin, Jongwoo et al. 2006). The application of the proposed neuro-fuzzy system in this thesis to the financial forecasting incorporating with reinforcement learning will be further investigated in the future work.

Appendix A: Experiment parameters configuration

θ is the firing strength threshold for the generation of initial rules (Section 3.1.1); λ is the overlap degree threshold for the merger of fuzzy membership functions (Section 3.1.2).

Data set	Chemical plant		Nonlinear system		Stock prediction	
Parameter	θ	λ	θ	λ	θ	λ
Noise-free	0.50	0.30	0.10	0.15	0.10	0.30
5% noise	0.35	0.50	0.10	0.10	0.10	0.55
10% noise	0.35	0.45	0.10	0.20	0.10	0.30
15% noise	0.80	0.35	0.10	0.10	0.10	0.25

Table A-1: Experiment parameter settings of the Nakanishi dataset.

Lane1	CV1		CV2		CV3	
Parameter	θ	λ	θ	λ	θ	λ
$\tau = 5$	0.10	0.30	0.40	0.30	0.10	0.30
$\tau = 15$	0.10	0.50	0.30	0.40	0.30	0.30
$\tau = 30$	0.30	0.40	0.40	0.40	0.20	0.30
$\tau = 45$	0.40	0.30	0.50	0.50	0.10	0.30
$\tau = 60$	0.30	0.40	0.10	0.60	0.20	0.30

Table A-2: Experiment parameter settings of traffic flow prediction data set: Lane 1.

Lane2	CV1		CV2		CV3	
Parameter	θ	λ	θ	λ	θ	λ
$\tau = 5$	0.10	0.30	0.30	0.30	0.30	0.30
$\tau = 15$	0.10	0.40	0.40	0.40	0.20	0.30
$\tau = 30$	0.20	0.30	0.10	0.30	0.40	0.50
$\tau = 45$	0.20	0.30	0.40	0.40	0.10	0.50
$\tau = 60$	0.30	0.40	0.50	0.30	0.20	0.60

Table A-3: Experiment parameter settings of traffic flow prediction data set: Lane 2.

Lane3	CV1		CV2		CV3	
Parameter	θ	λ	θ	λ	θ	λ
$\tau = 5$	0.10	0.30	0.30	0.40	0.10	0.30
$\tau = 15$	0.30	0.30	0.20	0.30	0.40	0.30
$\tau = 30$	0.30	0.30	0.30	0.30	0.20	0.30
$\tau = 45$	0.20	0.30	0.40	0.40	0.30	0.50
$\tau = 60$	0.20	0.30	0.50	0.40	0.50	0.50

Table A-4: Experiment parameter settings of traffic flow prediction data set: Lane 3.

Parameter	*Pat*	Two-Spiral	Iris: CV1	Iris: CV2	Iris: CV3
θ	0.30	0.10	0.50	0.85	0.20
λ	0.50	0.40	0.27	0.71	0.29

Table A-5: Experiment parameter settings of Pat Synthetic Pattern classification, Two-Spiral classification and Iris classification problems.

Parameter	CV1	CV2	CV3	CV4	CV5
θ	0.03	0.02	0.26	0.11	0.11
λ	0.05	0.07	0.08	0.19	0.05
Parameter	CV6	CV7	CV8	CV9	CV10
θ	0.11	0.12	0.01	0.02	0.05
λ	0.01	0.95	0.01	0.12	0.06

Table A-6: Experiment parameter settings of Pima Indian Diabetes dataset.

Parameter	CV1	CV2	CV3	CV4	CV5
θ	0.12	0.01	0.01	0.01	0.01
λ	0.06	0.01	0.01	0.61	0.01
Parameter	CV6	CV7	CV8	CV9	CV10
θ	0.01	0.01	0.01	0.01	0.01
λ	0.26	0.01	0.24	0.07	0.01

Table A-7: Experiment parameter settings of Urban Water Treatment dataset.

Parameter	CV1	CV2	CV3
θ	0.10	0.10	0.10
λ	0.16	0.17	0.37

Table A-8: Experiment parameter settings of Sonar dataset.

Parameter	CV1	CV2	CV3
θ	0.10	0.01	0.01
λ	0.03	0.17	0.33

Table A-9: Experiment parameter settings of Ovarian cancer dataset.

Parameter	CV1	CV2	CV3	CV4	CV5
θ	0.05	0.05	0.05	0.05	0.05
λ	0.95	0.55	0.25	0.60	0.50

Table A-10: Experiment parameter settings of Central Nervous System dataset.

Parameter	CV1	CV2	CV3
θ	0.90	0.15	0.10
λ	0.40	0.10	0.10

Table A-11: Experiment parameter settings of ICU ventilator data.

Parameter	CV1	CV2	CV3	CV4	CV5
θ	0.21	0.03	0.28	0.09	0.48
λ	0.92	0.90	0.83	0.90	0.89

Table A-12: Experiment parameter settings of Bank Failure Prediction dataset: Scenario 1.

Parameter	CV1	CV2	CV3	CV4	CV5
θ	0.03	0.37	0.53	0.02	0.01
λ	0.75	0.94	0.94	0.94	0.70

Table A-13: Experiment parameter settings of Bank Failure Prediction dataset: Scenario 2.

Parameter	CV1	CV2	CV3	CV4	CV5
θ	0.04	0.01	0.82	0.01	0.45
λ	0.70	0.95	0.67	0.57	0.53

Table A-14: Experiment parameter settings of Bank Failure Prediction dataset: Scenario 3.

Author's publications

International Journals:

Liu, F. G. S. Ng and C. Quek (2007). "RLDDE: A novel reinforcement learning based dimension and delay estimator for neural networks in time series prediction." Neurocomputing, **70(7-9)**: 1331-1341.

Liu, F., C. Quek and G. S. Ng (2007). "A novel generic Hebbian ordering based fuzzy rule base reduction approach to Mamdani neuro-fuzzy system." Neural Computation, **19(6)**:1658-1680.

International Conferences:

Liu, F., C. Quek and G. S. Ng (2005). "Input dimension estimation for neural network in time series prediction through reinforcement." The 12th International Conference on Neural Information Processing, Taipei, Taiwan, pp. 468-473.

Liu, F., C. Quek and G. S. Ng (2005). "Neural network model for time series prediction by reinforcement learning." International Joint Conference on Neural Networks, Montreal, Canada, pp. 809-814

Liu, F., C. Quek, G. S. Ng and Loh T. F. (2006). "Artificial ventilation modeling using neuro-fuzzy hybrid system." International Joint Conference on Neural Networks, Vancouver, BC, Canada, pp. 5166-5171.

References

Abe, S. and M.-S. Lan (1995). "A method for fuzzy rules extraction directly from numerical data and its application to pattern classification." IEEE Transactions on Fuzzy Systems **3**(1): 18-28.

Almuallim, H. and T. G. Dietterich (1994). "Learning Boolean concepts in the presence of many irrelevant features." Artificial Intelligence **69**: 279-306.

Anderson, J. R., D. Bothell and M. D. Byrne (2004). "An integrated theory of the mind." Psychological Review **111**(4): 1036-1060.

Ang, K. K. and C. Quek (2005). "RSPOP: rough set-based pseudo outer-product fuzzy rule identification algorithm." Neural Computation **17**(1): 205-243.

Ang, K. K. and C. Quek (2006). "Stock trading using RSPOP: A novel rough set-based neuro-fuzzy approach." IEEE Transactions on Neural Networks **17**(5): 1301-1315.

Ang, K. K., C. Quek and M. Pasquier (2003). "POPFNN-CRI(S): Pseudo outer product-based fuzzy neural network using the compositional rule of inference and singleton fuzzifier." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **33**(6): 838-849.

Banerjee, M., S. Mitra and S. K. Pal (1998). "Rough fuzzy MLP: Knowledge encoding and classification." IEEE Transactions on Neural Networks **9**(6): 1203-1216.

Baraldi, A. and P. Blonda (1999). "A survey of fuzzy clustering algorithms for pattern recognition - Part I." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **29**(6): 778-785.

Baraldi, A. and P. Blonda (1999). "A survey of fuzzy clustering algorithms for pattern recognition - Part II." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **29**(6): 786-801.

Bastian, A. (1994). "How to handle the flexibility of linguistic variables with applications." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **2**(4): 463-484.

Bernd, T. and L. Personnaz (1999). "Nonlinear black box modelling - fuzzy networks versus neural networks." Neural Computing & Applications **8**(2): 151-162.

Bouchon-Meunier, B. and Y. Jia (1992). "Linguistic modifiers and imprecise categories." International Journal of Intelligent Systems **7**: 25-36.

Broomhead, D. S. and D. Lowe (1988). "Multivariable functional interpolation and adaptive networks." Complex systems **2**: 321-355.

Bulter, R., S. P. Keenan, K. Inman, J., W. J. Sibbald and G. Block (1999). "Is there a preferred technique for weaning the difficult-to-wean patient? A systematic review of the literature." Critical Care Medicine **27**(11): 2331-2336.

Carpenter, G. A., S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen (1992). "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps." IEEE Transactions on Neural Networks **3**(5): 698-713.

Casillas, J., O. Cordon, M. J. Del Jesus and F. Herrera (2001). "Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems." Information Sciences **136**(1-4): 135-157.

Casillas, J., O. Cordon and F. Herrera (2002). "COR: a methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **32**(4): 526-537.

Casillas, J., O. Cordon, F. Herrera and L. Magdalena (2003). Accuracy improvements to find the balance interpretability-accuracy in linguistic fuzzy modeling: an overview. Accuracy improvements in linguistic fuzzy modeling. L. Magdalena, Springer**:** 3-24.

Casillas, J., O. Cordon, F. Herrera and L. Magdalena (2003). Interpretability improvements to find the balance interpretability-accuracy in fuzzy modeling: an overview. Interpretability issues in fuzzy modeling. L. Magdalena, Springer**:** 3-22.

Casillas, J., O. Cordon, F. Herrera and L. Magdalena (2003). Interpretability issues in fuzzy modeling, Springer.

Castro, J. L., J. J. Castro-Schez and J. M. Zurita (1999). "Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems." Fuzzy Sets and Systems **101**(3): 331-342.

Chakraborty, D. and N. R. Pal (2004). "A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification." IEEE Transactions on Neural Networks **15**(1): 110-123.

Cheng, P. Y. K. (2002). Predicting bank failures: A comparison of the cox proportional hazards model and the time varying covariates model. Nanyang Business School. Singapore, Nanyang Technological University.

Chow, T. W. S. and D. Huang (2005). "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information." IEEE Transactions on Neural Networks **16**(1): 213-224.

Cole, R. and J. Gunther (1995). "Separating the likelihood and timing of bank failure." Journal of Banking and Finance **19**: 1073-1089.

Combs, W. E. and J. E. Andrews (1998). "Combinatorial rule explosion eliminated by a fuzzy rule configuration." IEEE Transactions on Fuzzy Systems **6**(1): 1-11.

Cordon, O. and F. Herrera (2000). "A proposal for improving the accuracy of linguistic modeling." IEEE Transactions on Fuzzy Systems **8**(3): 335-344.

Cordon, O., F. Herrera, F. Hoffmann and L. Magdalena (2001). Genetic fuzzy systems: Evolutionary tuning and learning of fuzzy knowledge bases, World Scientific.

Cordon, O., F. Herrera and I. Zwir (2002). "Linguistic modeling by hierarchical systems of linguistic rules." IEEE Transactions on Fuzzy Systems **10**(1): 2-20.

Dempster, M. A. H. and V. Leemans (2006). "An automated FX trading system using adaptive reinforcement learning." Expert Systems with Applications **30**(3): 543-552.

Dojat, M., F. Pachet, Z. Guessoum, D. Touchard, A. Harf and L. Brochard (1997). "NeoGanesh: A working system for the automated control of assisted ventilation in ICUs." Artificial Intelligence in Medicine **11**(2): 97-117.

Er, M., J. and S. Wu (2002). "A fast learning algorithm for parsimonious fuzzy neural systems." Fuzzy Sets and Systems **126**(3): 337-351.

Espinosa, J. and J. Vandewalle (2000). "Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm." IEEE Transactions on Fuzzy Systems **8**(5): 591-600.

Fagan, L. M. (1980). VM: Representing time-dependent relations in a medical setting. Department of Computer Science, Stanford University. **Ph.D. Thesis**.

Fagan, L. M., J. C. Kunz and E. A. Feigenbaum (1979). Representation of dynamic clinical knowledge: Measurement interpretation in the intensive care unit. International Joint Conference on Artificial Intelligence, pp. 260-262.

Frank, H., K. Mario, K. Frank and R. Rajkumar (2005). Soft Computing: Methodologies And Applications, Springer.

Friedman, N., D. Geiger and M. Goldszmidt (1997). "Bayesian network classifiers." Machine Learning **29**(2-3): 131-163.

Furuhashi, T., T. Hasegawa, S. Horikawa and Y. Uchikawa (1993). An adaptive fuzzy controller using fuzzy neural networks. Fifth IFSA World Congress, pp. 769-772.

Gerard, D. (2005). Neural Networks: Methodology And Applications, Springer.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning, Addison-Wesley.

Gonzalez, A. and R. Perez (1998). "Completeness and consistency condictions for learning fuzzy rules." Fuzzy Sets and Systems **96**(1): 37-51.

Gonzalez, A. and R. Perez (2001). "Selection of relevant features in a fuzzy genetic learning algorithm." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **31**(3): 417-425.

Goode, K. M. (1993). Phase I respiratory model development: the model development: the blood-gas model outline and implementation using the SIMULINK and MATLAB simulation platforms, Department of Automatic Control and Systems Engineering, University of Sheffield and Department of Bioengineering, Hull Hospital NHS Trusts.

Goode, K. M., D. A. Linkens, P. R. Bourne and J. G. Cundill (1998). "Development of a fuzzy rule-based advisor for the maintenance of mechanically ventilated patients in ICU: a model-based approach." Biomedical Engineering: Applied Basis Communication **10**: 236-246.

Gorman, R. P. and T. J. Sejnowski (1988). "Analysis of hidden units in a layered network trained to classify sonar targets." Neural Networks **1**(1): 75-89.

Grossberg, S. (1980). "How does a brain build a cognitive code?" Psychological Review **87**: 1-51.

Haykin, S. (1999). Neural Networks: A Comprehensive Foundation, Prentice Hall.

Hebb, D. O. (1949). The organization of behavior. New York, John Wiley and Sons.

Hernandez-Sande, C., V. Moret-Bonillo and A. Alonso-Betanzos (1989). "ESTER: An expert system for management of respiratory Weaning Therapy." IEEE Transactions on Biomedical Engineering **36**: 551-564.

Herrera, F. and L. Martinez (2000). "A 2-tuple fuzzy linguistic representation model for computing with words." IEEE Transactions on Fuzzy Systems **8**(6): 746-752.

Hong, T. P. and C. Y. Lee (1999). "Finding relevant attributes and membership functions." Fuzzy Sets and Systems **103**(3): 139-151.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, pp. 2554-2558.

Huang, C.-J. and H.-C. Lin (2006). "Association between Adrenal Insufficiency and Ventilator Weaning." American Journal of Respiratory and Critical Care Medicine **173**: 276-280.

Ishibuchi, H., K. Murata and I. B. Turksen (1997). "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems." Fuzzy Sets and Systems

89(2): 135-150.

Ishibuchi, H. and T. Nakashima (2001). "Effect of rule weights in fuzzy rule-based classification systems." IEEE Transactions on Fuzzy Systems **9**(4): 506-515.

Ishibuchi, H., K. Nozaki, N. Yamamoto and H. Tanaka (1995). "Selecting fuzzy if-then rules for classification problems using genetic algorithms." IEEE Transactions on Fuzzy Systems **3**(3): 260-270.

James, J. B. (2005). Simulating Fuzzy Systems, Springer.

Jang, J.-S. R. (1993). "ANFIS: Adaptive-network-based fuzzy inference system." IEEE Transactions on Systems, Man and Cybernetics **23**(3): 665-685.

Jangmin, O., L. Jongwoo, W. L. Jae and T. Z. Byoung (2006). "Adaptive stock trading with dynamic asset allocation using reinforcement learning." Information Sciences **176**(15): 2121-2147.

Jensen, R. and Q. Shen (2004). "Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches." IEEE Transactions on Knowledge and Data Engineering **16**(12): 1457-1471.

Jin, Y. (2000). "Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement." IEEE Transactions on Fuzzy Systems **8**(2): 212-221.

John, G. H., R. Kohavi and K. Pfleger (1994). Irrelevant features and the subset selection problem. 11th International Conference on Machine Learning, New Brunswick. pp. 121-129.

Juang, C. F. (2002). "A TSK-type recurrent fuzzy network for dynamic systems processing by neural network and genetic algorithms." IEEE Transactions on Fuzzy Systems **10**(2): 155-170.

Juang, C. F. and C. T. Lin (1998). "An online self-constructing neural fuzzy inference network and its applications." IEEE Transactions on Fuzzy Systems **6**(1): 12-32.

Juang, C. F. and C. T. Lin (1999). "A recurrent self-organizing neural fuzzy inference network." IEEE Transactions on Neural Networks **10**(4): 828-845.

Kasabov, N. (1996). Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering. Cambridge, MA., MIT Press.

Kasabov, N. (2001). "Evolving fuzzy neural networks for supervised/unsupervised online knoledge-based learning." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **31**(6): 902-918.

Kasabov, N. (2003). Evolving connectionist systems: methods and applications in bioinformatics, brain study and intelligent machines, Springer.

Kasabov, N. K. (2001). "Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **31**(6): 902-918.

Kasabov, N. K. and Q. Song (2002). "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction." IEEE Transactions on Fuzzy Systems **10**(2): 144-154.

Kaya, M. and R. Alhaji (2006). "Utilizing genetic algorithms to optimize membership functions for fuzzy weighted association rules mining." Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies **24**(1): 7-15.

Kim, J. and N. Kasabov (1999). "HyFIS: Adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems." Neural Networks **12**(9): 1301-1319.

Kira, K. and L. Rendell (1992). A practical approach to feature selection. 9th International Conference on Machine Learning, Aberdeen, Scotland, pp. 249-256.

Klose, A., A. Nurnberger and D. Nauck (1998). Some approaches to improve the interpretability of neuro-fuzzy classifiers. The 6th European Congress on Intelligent Techniques and Soft Computing, Aachen, Germany, pp. 629-633.

Kohavi, R. and G. H. John (1997). "Wrappers for feature subset selection." Artificial Intelligence **97**(1-2): 273-324.

Kohonen, T. K. (1982). "Self-organized formation of topologically correct feature maps." Biological Cybernetics **43**(1): 59-69.

Krone, A. and H. Kiendl (1994). Automatic generation of positive and negative rules for two-way fuzzy controllers. The Second European Congress on Intelligent Techniques and Soft Computing, Aachen Germany, pp. 438-447.

Kwak, N. and C. H. Choi (2002). "Input feature selection by mutual information based on Parzen window." IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(12): 1667-1671.

Kwok, H. F., D. A. Linkens, M. Mahfouf and G. H. Mills (2003). "Rule-base derivation for intensive care ventilator control using ANFIS." Artificial Intelligence in Medicine **29**(3): 185-201.

Kwok, H. F., D. A. Linkens, M. Mahfouf and G. H. Mills (2004). "SIVA: A hybrid knowledge-and-model-based advisory system for intensive care ventilators." IEEE Transactions on Information Technology in Biomedicine **8**(2): 161-172.

Lane, W., S. Looney and J. Wansley (1986). "An application of the Cox proportional hazards model to bank failure." Journal of Banking and Finance **10**: 511-531.

Lang, K. J. and Witbrock (1988). Learning to tell two spirals apart. Proceedings of 1989 Connectionist Models Summer School, pp. 52-59.

Lee, C. H. and C. C. Teng (2001). "Tuning of membership functions for fuzzy neural systems." Asian Journal of Control **3**(3): 216-225.

Lee, H. M., C. M. Chen, J. M. Chen and Y. L. Jou (2001). "An efficient fuzzy classifier with feature selection based on fuzzy entropy." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **31**(3): 426-432.

Lee, H. M., J. M. Chen and C. L. Liu (2002). "Inconsistency resolution and rule insertion for fuzzy rule-based systems." Journal of Information Science and Engineering **18**(2): 187-210.

Li, Y., C. Campbell and M. Tipping (2002). "Bayesian automatic relevance determination algorithms for classifying gene expression data." Bioinformatics **18**(10): 1332-1339.

Lin, C. T. and C. S. G. Lee (1991). "Neural-network-based fuzzy logic control and decision system." IEEE Transactions on Computers **40**(12): 1320-1336.

Lin, J. (1996). "Integration of weighted knowledge bases." Artificial Intelligence **83**: 363-378.

Lin, J. (1996). "A semantics for reasoning consistently in the presence of inconsistency." Artificial Intelligence **86**: 75-95.

Liu, F., C. Quek and G. S. Ng (2007). "A novel generic Hebbian ordering based fuzzy rule base reduction approach to Mamdani neuro-fuzzy system." Neural Computation **19**(6): 1658-1680.

Liu, F., C. Quek, G. S. Ng and T. F. Loh (2006). Artificial ventilation modeling using neuro-fuzzy hybrid system. International Joint Conference on Neural Networks, Vancouver, BC, Canada, pp. 5166-5171.

Lozowski, A., T. J. Cholewo and J. M. Zurada (1996). Crisp rule extraction from perceptron network classifiers. International Conference on Neural Networks, Washington, D. C., pp. 94-99.

Mamdani, E. H. and S. Assilian (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." International Joural of Human Machine Studies **7**(1): 1-13.

Martin, D. (1977). "Early warning of bank failure: A logit regression approach." Journal of Banking and Finance **1**(3): 249-276.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." Bulletin of Mathematical Biophysics **5**: 115-133.

Miller, G. A., E. Galanter and K. H. Pribam (1960). Plans and Structure of Behaviour. New York, Holt, Rinehart and Winston.

Minsky, M. L. (1954). Theory of neural-analog reinforcement systems and its application to the brain-model problem. Princeton, NJ, Princeton University. **Ph.D. Thesis**.

Minsky, M. L. (1961). Steps towards artificial intelligence. Proceedings of the Institute of Radio Engineers, pp. 8-30.

Minsky, M. L. and S. A. Papert (1969). Perceptrons. Cambridge, MA, MIT Press.

Mitra, S. and Y. Hayashi (2000). "Neuro-fuzzy rule generation: Survey in soft computing framework." IEEE Transactions on Neural Networks **11**(3): 748-768.

Mitra, S., P. Mitra and S. K. Pal (2001). "Evolutionary modular design of rough knowledge-based network using fuzzy attributes." Neurocomputing **36**(1): 45-66.

Molina, L. C., L. Belanche and A. Nebot (2002). Feature selection algorithms: A survey and experimental evaluation. 2002 IEEE International Conference on Data Mining, Maebashi City, Japan, pp. 306-313.

Moody, J. and M. Saffell (2001). "Learning to trade via direct reinforcement." IEEE Transactions on Neural Networks **12**(4): 875-889.

Moody, J., L. Wu, Y. Liao and M. Saffell (1998). "Performance functions and reinforcement learning for trading systems and portfolios." Journal of Forecasting **17**(5-6): 441-470.

Nakanishi, H., I. B. Turksen and M. Sugeno (1993). "A review and comparison of six reasoning methods." Fuzzy Sets and Systems **57**(3): 257-294.

Nauck, D., F. Klawonn and R. Kruse (1997). Foundations of Neuro-Fuzzy Systems. New York, John Wiley.

Nemoto, T., G. E. Hatzakis, C. W. Thorpe, R. Olivenstein, S. Dial and J. H. T. Bates (1999). "Automatic control of pressure support mechanical ventilation using fuzzy logic." American Journal of Respiratory and Critical Care Medicine **160**: 550-556.

Ng, G. S., C. Quek and H. Jiang (2008). "FCMAC-EWS: A bank failure early warning system based on a novel localized pattern learning and semantically associative fuzzy neural network." Expert Systems with Applications **34**(2): 989-1003.

Nguyen, M. N., D. Shi and C. Quek (2006). "FCMAC-BYY: Fuzzy CMAC using bayesian Ying-Yang learning." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **36**(5): 1180-1190.

O'Reilly, R. C. and Y. Munakata (2000). Computational Explorations in Cognitive Neuroscience. London, England, MIT Press.

Paiva, R. P. and A. Dourado (2004). "Interpretability and learning in neuro-fuzzy systems." Fuzzy Sets and Systems **147**(1): 17-38.

Pal, N. R. and K. Pal (1999). "Handling of inconsistent rules with an extended model of fuzzy reasoning." Journal of Intelligent and Fuzzy Systems **7**: 55-73.

Pal, S. K., S. Mitra and P. Mitra (2003). "Rough Fuzzy MLP: Modular evolution, rule generation and evaluation." IEEE Transactions on Knowledge and Data Engineering **15**(1): 14-25.

Papadelis, C., N. Maglaveras, C. Kourtidou-Papadeli, P. Bamidis, M. Albani, K. Chatzinikolaou and K. Pappas (2006). "Quantitative multichannel EEG measure predicting the optimal weaning from ventilator in ICU patients with acute respiratory failure." Clinical Neurophysiology **117**(4): 752-770.

Pawlak, Z. (1991). Rough sets: Theoretical aspects of reasoning about data. Boston, Kluwer.

Pawlak, Z. and A. Skowron (2007). "Rudiments of rough sets." Information Sciences **177**(1): 3-27.

Peng, H., F. Long and C. Ding (2005). "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(8): 1226-1238.

Polaschek, J. X., G. W. Rutledge, S. K. Andersen and L. M. Fagan (1993). "Using belief networks to interpret qualitative data in the ICU." Respiratory Care **38**(1): 60-71.

Pomeroy, S. L., P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. H. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander and T. R. Golub (2002). "Prediction of central nervous system embryonal tumour outcome based on gene expression." Nature **415**(6870): 436-442.

Quek, C. and A. Singh (2005). "POP-Yager: A novel self-organizing fuzzy neural network based on the Yager inference." Expert Systems with Applications **29**(1): 229-242.

Quek, C. and R. W. Zhou (1999). "POPFNN-AARS(S): A pseudo outer-product based fuzzy neural network." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **29**(6): 859-870.

Quek, C. and R. W. Zhou (2006). "Structure and learning algorithms of a nonsingleton input fuzzy neural network based on the approximate analogical reasoning schema." Fuzzy Sets and Systems **157**(13): 1814-1831.

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. San Mateo, Morgan Kaufmann.

Rosenblatt, F. (1958). "The Perceptron: A probabilistic model for information storage and organization in the brain." Psychological Review **65**: 386-408.

Rudowski, R., C. Frostell and H. Gill (1989). "A knowledge-based support system for mechanical ventilation of lungs. The KUSIVAR concept and prototype." Computer Methods Programs Biomedicine **30**(1): 59-70.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). "Learning internal representations by error propagation." Parallel Distributed Processing **1**(8): 318-362.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). "Learning representations of back-propagation errors." Nature **323**: 533-536.

Rumelhart, D. E. and J. L. McClelland (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA, MIT Press.

Rutkowski, L. and K. Cpalka (2003). "Flexible neuro-fuzzy systems." IEEE Transactions on Neural Networks **14**(3): 554-574.

Rutkowski, L. and K. Cpalka (2005). "Designing and learning of adjustable quasi-triangular norms with applications to neuro-fuzzy systems." IEEE Transactions on Fuzzy Systems **13**(1): 140-151.

Rutledge, G. W. (1994). VentSim: A simulation model of cardiopulmonary physiology. Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care, pp. 878-883.

Rutledge, G. W., S. K. Andersen, J. X. Polaschek and L. M. Fagan (1990). A belief network model for interpretation of ICU data. Proceedings of the Fourteenth Annual Symposium of Computer Applications in Medical Care, pp. 828-832.

Rutledge, G. W., G. Thomsen, I. Beinlich, B. Farr, M. Kahn, L. Sheiner and L. M. Fagan (1990). VentPlan: An architecture for combining qualitative and quantitative computation. AAAI Spring Symposium Series AI in Medicine, pp. 159-160.

Rutledge, G. W., G. Thomsen, I. Beinlich, B. Farr, L. Sheiner and L. M. Fagan (1989). Combining qualitative and quantitative computation in a ventilator therapy planner. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care, pp. 315-319.

Rutledge, G. W., G. Thomsen, B. Farr, M. Tovar, L. Sheiner and L. M. Fagan (1991). VentPlan: A ventilator-management advisor. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical care, pp. 869-871.

Schuh, C., M. Hiesmayr, K. P. Adlassnig, C. Zelenka and E. P. Klement (1998). Integration of crisp and fuzzy controlled weaning in an ICU PDMS. Proceedings of the world automation Congress, WAC'98, pp. 299-304.

Schuh, C., C. Zelenka and M. Hiesmayr (2000). FuzzyKBWean: A computer program for fuzzy knowledge-based weaning of ICU patients. Proceedings of the ERUDIT-Workshop, pp. 38-44.

Schummer, M., W. V. Ng, R. E. Bumgarner, P. S. Nelson, B. Schummer, D. W. Bednarski, L. Hassell, R. L. Baldwin, B. Y. Karlan and L. Hood (1999). "Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas." Gene **238**(2): 375-385.

Seising, R., C. Schuh and K. P. Adlassnig (2003). Medical knowledge, fuzzy sets and expert system. Workshop on Intelligent and Adaptive Systems in Medicine, Prague.

Setnes, M. and R. Babuska (2001). "Rule base reduction: Some comments on the use of orthogonal transforms." IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews) **31**(2): 199-206.

Setnes, M., R. Babuska, U. Kaymak and H. R. van Nauta Lemke (1998). "Similarity measures in fuzzy rule base simplification." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **28**(3): 376-386.

Setnes, M., R. Babuska and H. B. Verbruggen (1998). "Complexity reduction in fuzzy modeling." Mathematics and Computers in Simulation **46**(5-6): 509-518.

Setnes, M., J. Casillas, O. Cordon and M. J. Del Jesus (2002). "Some relationships between fuzzy and random set-based classifiers and models." International Journal of Approximate Reasoning **29**(2): 175-213.

Shahsavar, N., H. Gill, U. Ludwigs, A. Carstensen, H. Larsson, O. Wigertz and G. Matell (1994). "VentEx: An on-line knowledge-based system to support ventilator management." Technology and Health Care **1**: 233-243.

Shahsavar, N., H. Gill, O. Wigertz, C. Frostell, G. Matell and U. Ludwigs (1991). "KAVE: A tool for knowledge acquisition to support artificial ventilation." International Jounal of Computer Methods and Programs in Biomedicine **34**: 115-123.

Shen, Q. and A. Chouchoulas (2002). "A fuzzy-rough approach for generating classification rules." Pattern Recognition **35**(11): 341-354.

Shen, Q. and R. Jensen (2004). "Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring." Pattern Recognition **37**(7): 1351-1363.

Silipo, R. and M. Berthold (2000). "Input features' impact on fuzzy decision processes." IEEE

Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **30**(6): 821-834.

Sinkey, J. J. (1975). "A multivariate statistical analysis of the characteristics of problem banks." Journal of Finance **1**: 21-36.

Song, Q. and N. Kasabov (2005). "NFI: A neuro-fuzzy inference method for transductive reasoning." IEEE Transactions on Fuzzy Systems **20**(2-3): 128-158.

Stamou, G. B. and S. G. Tzafestas (1999). "Fuzzy relation equations and fuzzy inference systems: An inside approach." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **29**(6): 694-702.

Sugeno, M. and G. T. Kang (1988). "Structure identification of fuzzy model." Fuzzy Sets and Systems **28**(1): 15-33.

Sugeno, M. and T. Yasukawa (1993). "A fuzzy-logic-based approach to qualitative modeling." IEEE Transactions on Fuzzy Systems **1**(1): 7-31.

Sun, Y., I. Kohane and A. R. Stark (1994). Fuzzy logic assisted control of inspired oxygen in ventilated newborn infants. Proceedings of Annual Symposium on Computer Applications in Medical Care, pp. 756-761.

Takagi, T. and M. Sugeno (1985). "Fuzzy identification of systems and its application to modeling and control." IEEE Transactions on Systems, Man and Cybernetics **15**(1): 116-132.

Tan, G. K. (1997). Feasibility of predicting congestion states with neural networks. School of Civil Structural Engineering. Singapore, Nanyang Technological University. **Final Year Project Report**.

Tan, T. Z., C. Quek and G. S. Ng (2005). "Ovarian cancer diagnosis by hippocampus and neocortex-inspired learning memory structures." Neural Networks **18**(5): 818-825.

Thomsen, G. and L. Sheiner (1989). SIMV: An application of mathematical modeling in ventilator management. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care, pp. 320-324.

Tung, W. L. (2003). A generalized framework for fuzzy neural architecture. School of Computer Engineering. Singapore, Nanyang Technological University. **Ph.D. Thesis**.

Tung, W. L. and C. Quek (2002). "GenSoFNN: A generic self-organizing fuzzy neural network." IEEE Transactions on Neural Networks **13**(5): 1075-1086.

Tung, W. L., C. Quek and P. Cheng (2004). "GenSo-EWS: A novel neural-fuzzy based early warning systems for predicting bank failures." Neural Networks **17**(4): 567-587.

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory, Springer.

Wang, J. H. and J. Hao (2006). "A new version of 2-tuple fuzzy linguistic representation model for computing with words." IEEE Transactions on Fuzzy Systems **14**(3): 435-445.

Weavind, L., A. D. Shaw and T. W. Feeley (2000). "Monitoring ventilator weaning - Predictors of success." Journal of Clinical Monitoring and Computing **16**(5-6): 409-416.

Widrow, B. and M. E. Hoff (1960). "Adaptive switching circuits." IRE WESCON Convention Record: 96-104.

Wu, C. H. (1997). Knowledge validation/verification of rule-based expert systems. Taiwan, National Sun Yat-Sen University. **Ph. D. Thesis**.

Wu, S. and M. Er, J. (2000). "Dynamic fuzzy neural networks - a novel approach to function approximation." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **30**(2): 358-364.

Wu, S. and M. Er, J. (2001). "A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks." IEEE Transactions on Fuzzy Systems **9**(4): 578-594.

Xu, R. and D. Wunsch II (2005). "Survey of clustering algorithms." IEEE Transactions on Neural Networks **16**(3): 645-678.

Yamakawa, T., E. Uchino, T. Miki and H. Kusanagi (1992). A neo fuzzy neuron and its application to system identification and prediction of the system behaviour. 2nd International Conference on Fuzzy Logic and Neural Networks, pp. 477-483.

Yen, J. and L. X. Wang (1999). "Simplifying fuzzy rule-based models using orthogonal transformation methods." IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) **29**(1): 13-24.

Zadeh, L. A. (1972). "A fuzzy-set theoretic interpretation of linguistic hedges." Journal of Cybernetics **2**(2): 4-34.

Zadeh, L. A. (1973). "Outline of a new approach to the analysis of complex systems and decision processes." IEEE Transactions on Systems, Man and Cybernetics **3**: 28-44.

Zadeh, L. A. (1975). "The concept of a linguistic variable and its application to approximate reasoning. Parts I, II and III." Information Science **8,8,9**(199-249, 301-357, 43-80).

Zadeh, L. A. (1994). "Fuzzy logic, neural networks and soft computing." Communications of ACM **37**(3): 77-84.

Zadeh, L. A. (1996). "Fuzzy logic = computing with words." IEEE Transactions on Fuzzy Systems **4**(2):

103-111.

Zhou, R. W. and C. Quek (1996). "POPFNN: A pesudo outer-product based fuzzy neural network." Neural Networks **9**(9): 1569-1581.