A Framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS)

Chang Su Lee

B.S. Electronic Engineering

M.S. Electrical and Computer Engineering

This thesis is presented for the degree of

Doctor of Philosophy of Engineering of

The University of Western Australia

School of Electrical, Electronic and Computer Engineering

July 2009

Abstract

Fuzzy inference systems (FIS) are information processing systems using fuzzy logic mechanism to represent the human reasoning process and to make decisions based on uncertain, imprecise environments in our daily lives. Since the introduction of fuzzy set theory, fuzzy inference systems have been widely used mainly for system modeling, industrial plant control for a variety of practical applications, and also other decision-making purposes; advanced data analysis in medical research, risk management in business, stock market prediction in finance, data analysis in bioinformatics, and so on.

Many approaches have been proposed to address the issue of automatic generation of membership functions and rules with the corresponding subsequent adjustment of them towards more satisfactory system performance. Because one of the most important factors for building high quality of FIS is the generation of the knowledge base of it, which consists of membership functions, fuzzy rules, fuzzy logic operators and other components for fuzzy calculations. The design of FIS comes from either the experience of human experts in the corresponding field of research or input and output data observations collected from operations of systems. Therefore, it is crucial to generate high quality FIS from a highly reliable design scheme to model the desired system process best.

Furthermore, due to a lack of a learning property of fuzzy systems themselves most of the suggested schemes incorporate hybridization techniques towards better performance within a fuzzy system framework. A fuzzy system combined with neural networks is a representative example of hybrid fuzzy systems incorporated to learn the pattern of input and output relations so that the fuzzy systems trained can produce the output against new unknown input data. Other hybridization cases are, for instance, genetic fuzzy systems to optimize the corresponding objective function according to their system purposes, statistical function-combined fuzzy systems for modeling and analyzing huge data gathered for extracting useful information, and so on.

Even though most of these systems mentioned have provided very encouraging and satisfactory results to achieve their goals and solve problems, they have suffered from the computational complexity needed to calculate their system outputs. One problem

lies in the difficulties associated with the maximum number of resulting fuzzy rules, which increases exponentially when higher number of input features is employed. As a consequence, the computational load required to search for a corresponding fuzzy rule becomes very heavy. The fuzzy rules generated also need to be examined for their validity for use as appropriate fuzzy rules before carrying out the inference process. This validity checking process is to ensure a full coverage of the generated rules to represent the given knowledge. Another is that the initially obtained membership functions and rules based on *a prior* knowledge are often in need of advanced system adjustment and refinement towards higher accuracy. This systematic enhancement is required to update the FIS in order to produce flexible and robust fuzzy systems for unexpected unknown inputs from real-world environments.

This thesis proposes a general framework of Adaptive T-S (Takagi-Sugeno) type Rough-Fuzzy Inference Systems (ARFIS) for a variety of practical applications in order to resolve the problems mentioned above in the context of a Rough-Fuzzy hybridization scheme. Rough set theory is employed to effectively reduce the number of attributes that pertain to input variables and obtain a minimal set of decision rules based on input and output data sets. The generated rules are examined by checking their validity to use them as T-S type fuzzy rules. Using its excellent advantages in modeling non-linear systems, the T-S type fuzzy model is chosen to perform the fuzzy inference process. A T-S type fuzzy inference system is constructed by an automatic generation of membership functions and rules by the Fuzzy C-Means (FCM) clustering algorithm and the rough set approach, respectively. The generated T-S type rough-fuzzy inference system is then adjusted by the least-squares method and a conjugate gradient descent algorithm towards better performance within a fuzzy system framework.

To show the viability of the proposed framework of ARFIS, the performance of ARFIS is compared with other existing approaches in a variety of practical applications; pattern classification, face recognition, and mobile robot navigation. The results are very satisfactory and competitive, and suggest the ARFIS is a suitable new framework for fuzzy inference systems by showing a better system performance with less number of attributes and rules in each application.

Acknowledgements

Firstly, many thanks to my supervisors; Professor Thomas Bräunl for supervising my research and offering much advice to improve my study, robotics research, associated research environment and others, and to my co-supervisor Professor Anthony Zaknich for supervising theoretical developments of my research with much helpful advice to go further than my boundaries, and to the previous co-supervisor Dr Guilhereme DeSouza previously at UWA for suggesting the direction of my research at the early stage. Other thanks to the previous Head of the School, Professor Gary Bundell for allowing me to have teaching opportunities for years and for excellent consultations regarding my study at UWA.

Also thanks to my colleagues, James Ng and Adrian Boeing in the Robotics and Automation Laboratory at UWA for sharing with me their valuable feedback for my research, working together in teaching subjects and on some projects. To Pantazis Houlis from the Control Systems Laboratory at UWA for sharing valuable time for my research and time with my family as well, and Dr Martin Masek, a lecturer in the Computer Science Department at ECU, offered a very good support in many ways for my research and for my family as a family friend.

Finally, a very sincere thankyou message with my whole heart to my family; To my wife Sung-Joo Lee, my adorable son Joshua Young-Min Lee, families of sisters-in-law, and especially to my mother-in-law for all the support for my study and family life here in Perth, Australia. Another sincere thankyou message to my parents in Korea and my brother's family for all their support with daily prayers all the time for my family for many years.

Most importantly, thanks be to God who always is with me to leading me into his glory, plans, and his love in every daily affair. I have no doubts in the fact that I was, am, and will be in his good plans, love, blessings, and glory for the entirety of my life in the name of Jesus Christ.

Declaration

I declare that the research presented in this thesis is my own work unless otherwise specified. This work has not been submitted for any other degree or professional qualification except as specified.

Chang Su Lee

Table of Contents

Table of Figures

Table of Tables

Table of Acronyms

AFIS	Adaptive Fuzzy Inference Systems
AMFS	Adaptive Membership Function Scheme
ANOVA	Analysis of variance
ARFIS	Adaptive Rough-Fuzzy Inference Systems
EBPNN	Error-Back Propagation Neural Network
FCM	Fuzzy C-Means
FIS	Fuzzy Inference Systems
GA	Genetic Algorithm
GMP	Generalized Modus Ponens
I/O	Input-Output
ICA	Independent Component Analysis
LMS	Least Mean Squares
LSE	Least Square Estimate
MF	Membership Functions
MIMO	Multi-Input-Multi-Output
MISO	Multi-Input-Single-Output
ND	Non-Deterministic
PCA	Principal Component Analysis
PI	Performance Index
PID	Proportional-Integral-Derivative
RBF	Radial Basis Function
RFC	Rough-Fuzzy Controller
RLS	Recursive Least Squares
RMSE	Root-Mean-Square-Error
SISO	Single-Input-Single-Output
SRA	Similarity of the Rule Antecedent
SRC	Similarity of the Rule Consequent
T-S	Takagi-Sugeno

Chapter 1

1. Introduction

The decision-making process in human brains is one of the most important information processing routines that act in response, and according to the given input. Human beings obtain data from the external environment as a *Data Acquisition* process using their sensors, such as eyes, ears, skin, and so on. Then different sets of abstract categories are created related to the data gathered, which is called *Data Interpretation*. These categorized sets are classified according to previously accumulated experience of recognition and linked with pre-established or new concepts as a *Knowledge Base*. The interpreted knowledge of the given input is fed into the *Reasoning Mechanism* to produce the final decision. Finally, the decision output leads to an action for humans to react according to the given input. These procedures described are shown in Figure 1.1 as a series of modules for carrying out the decision-making process.

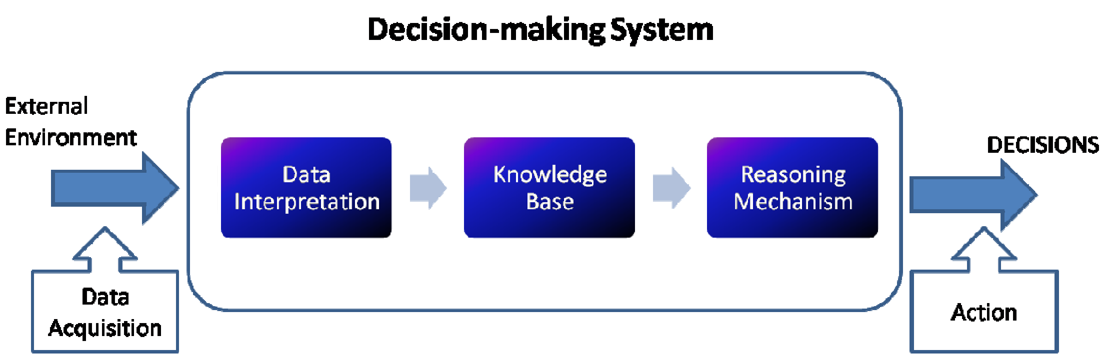


Figure 1.1 The decision–making process

1.1 Introduction and Main Goals

The fuzzy inference process is a type of information processing, which represents decision-making using fuzzy set theory about uncertainty, imprecision, and vagueness of objects that are of research interest. In fuzzy inference systems, the reasoning task is carried out based on fuzzy rules composed by fuzzy system developers using fuzzy linguistic variables assigned with the corresponding concept of the interpreted input data. The decision output of the reasoning is calculated by a fuzzy inference engine linked with other components in the knowledge-base of fuzzy inference systems (FIS).

The fuzzy inference system in Figure 1.2 shows the main process of fuzzy inference. When a given input is fed into the pre-designed fuzzy inference system based on *a priori* knowledge, each attribute of the input is fuzzified using membership functions in the knowledge-base. The fuzzified input data are fed into fuzzy inference rules to produce rule output for each rule. The rule outputs are calculated according to the design of fuzzy operators in fuzzy rules. Then a final output is calculated by de-fuzzification process. During the calculations of all fuzzy operations mentioned, the fuzzy logic operators stored in the knowledge-base of the FIS are employed.

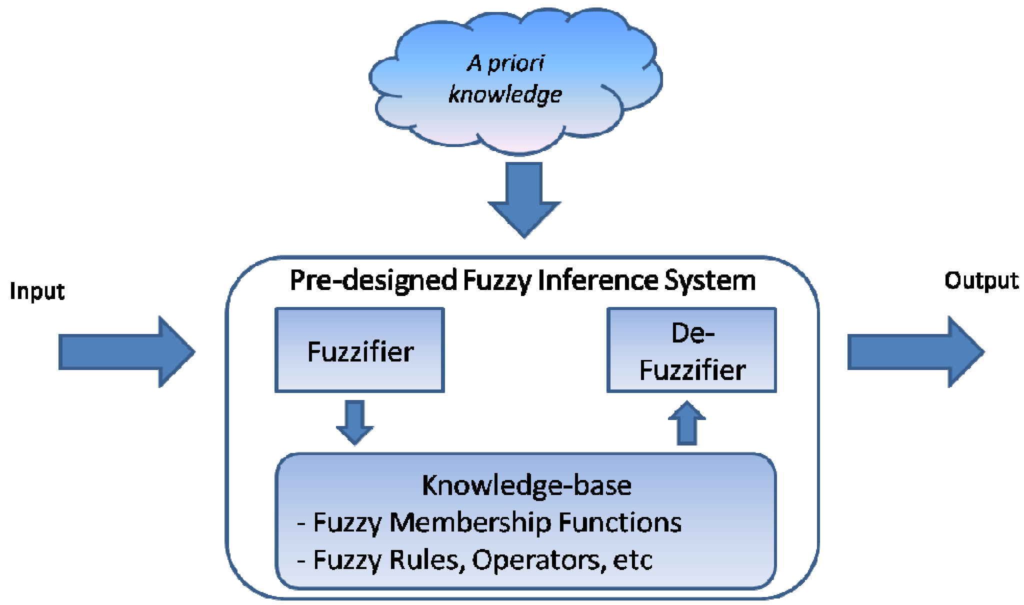


Figure 1.2 The fuzzy inference process

Since the first initiation of fuzzy logic by Zadeh [1], fuzzy inference systems have been developed to improve performance in the decision-making scheme over the past decades in; non-linear system modeling [2], [3], [4], industrial plant control [5], [6], robotics [7], [8], system identification [9], [10], [11], and so on. A number of approaches have been suggested to enhance the fuzzy logic-based decision-making mechanism in order to resolve its problems and handle issues related to fuzzy inference systems.

The main attention in the development of fuzzy inference systems has been focused on the automatic generation of membership functions and rules, and the hybrid techniques for providing the conventional fuzzy systems with a learning capability towards better system performance. One of the significant factors to assess the performance of fuzzy inference systems is the design of the knowledge-base, because the design scheme of a choice of fuzzy logic operators, a type of membership functions, and a composition strategy of fuzzy rules will produce a final output in order to approximate the given input. Every FIS has a different system purpose and details according to the objective of the system. The design of a FIS is determined by either the experience of experts in a particular field of the research or input and output data collection from system operations. Accordingly, the quality and the system performance of a FIS are dependent on the system design. In other words, it is critical to design a better quality FIS by a highly reliable system design scheme to achieve the desired system process.

Hybrid techniques in fuzzy inference systems have been proposed to supplement the existing fuzzy systems mainly with learning and optimizing capabilities due to the lack of flexibility of the conventional fuzzy inference system itself. A common hybrid system is a neural-fuzzy system which is a combination of neural networks and fuzzy systems [12], [13], [14], [15]. For other examples, there are GA(Genetic Algorithm)-fuzzy systems to optimize their objective function using GA according to their system purposes [16], [17], [18], and statistical function-combined fuzzy systems [19], [20] to analyze extremely high dimensional data sets or to model complex systems based on massive data observations to extract information of research interest.

However, most of the fuzzy inference systems mentioned above have been hampered by the computational complexity to calculate the final output of their systems. Once a higher number of input features is used, the maximum number of resulting fuzzy rules is increased exponentially. Accordingly, the computational burden for fuzzy systems to search for a corresponding fuzzy rule to fire is extremely heavy. This difficulty may cause system delay or even system malfunction. Also the generated knowledge-base should be investigated for its validity for use at the stage of the system design. In particular, the obtained fuzzy rules should be examined to ensure the full coverage of the input and output relation of the given information. Moreover, the initial system design of FIS should be enhanced through a system evaluation process towards higher system performance and better robustness against all the possible unpredicted inputs from the real-world environment.

The major contribution of this thesis is a development of a framework of Adaptive T-S (Takagi-Sugeno) type Rough-Fuzzy Inference Systems (ARFIS) to solve the problems described and to handle issues mentioned in the context of a Rough-Fuzzy hybridization scheme. Rough set theory [36] here is utilized to reduce the number of features using the dependency of attributes, and to generate rules using decision rule generation. An efficient knowledge-reduction is carried out according to the proposed rough set approach. The T-S type fuzzy model [7] is chosen to perform fuzzy inference utilizing its advantages as a universal approximator. The membership functions and fuzzy rules in the knowledge-base are generated using the Fuzzy C-Means (FCM) clustering method [22] and the rough set approach, respectively. After the rule generation, the rules are examined for validity and their suitability as T-S type fuzzy inference rules to ensure the full coverage of the input and output relation of the given knowledge. Once the whole system is established, the system performance evaluation is done based on the Root-Mean-Square-Error (RMSE) measure between the desired target output and the actual current output. If the RMSE measure is not satisfactory, the adjustment of membership functions and the rule refinement procedure is activated towards better system performance.

The proposed system was applied to a variety of applications to show the viability of the proposed framework of ARFIS. Results shown in experimental evaluation section are highly competitive, and suggest that the ARFIS is a suitable new framework for rough-fuzzy inference systems.

1.2 Thesis Outline

The rest of this thesis is organized as follows:

● **Chapter 1:** *Introduction to this research*. The decision-making process of the human brain, fuzzy inference systems, the system design scheme, problems of FIS and the existing approaches, and the main contribution of this thesis are briefly described.

● **Chapter 2:** *Background of the T-S type fuzzy model*. The T-S type fuzzy model is reviewed giving its basic definition and theory of a general T-S fuzzy model as a universal approximator. In this chapter, the focus is on the examination of the T-S type fuzzy model as an excellent tool representing the universal approximator to show its advantages in non-linear system modeling.

● **Chapter 3:** *Review of Rough Set theory*. This chapter presents the review of most of the functionalities of rough set theory; from the definition of rough sets to decision rule generation and knowledge-reduction methods. Most of the functionalities in rough set theory are described via examples, since it is the best way to explain rough set theory for readers.

● **Chapter 4:** *Theoretical modeling of Rough-Fuzzy hybridization*. A theoretical investigation of rough sets, fuzzy sets, and rough-fuzzy sets is mentioned and focused on the combination of rough sets and fuzzy sets. This hybrid technique is explained from the perspective of α-level sets as a tool for set analysis. Many proposals on the combination of fuzzy and rough sets are reviewed.

● **Chapter 5:** *Development of a framework of Adaptive T-S type Rough–Fuzzy Inference Systems (ARFIS)*. This is the main contribution of this thesis. This chapter describes all the modules of the proposed system to build a framework of ARFIS using the automatic generation of the T-S type fuzzy model and the efficient knowledge-reduction method.

● **Chapter 6:** *Applications to Pattern Classification, Face Recognition, and Mobile Robot Navigation*. The proposed framework of ARFIS was applied to each application to obtain better system performance with less number of input features and rules. As a result, the proposed approach has produced better results in each application with the absolute minimal size of the given knowledge without losing its original essential information. It has been shown that the system performance of the proposed framework of ARFIS is very encouraging, satisfactory, and competitive.

● **Chapter 7:** *Conclusion and Discussion*. The advantages and the potentials of this research are discussed. From the perspective of rough-fuzzy hybridization scheme, a variety of issues are mentioned about theoretical and practical aspects of this research including the future direction in this field of research. This chapter also describes the future work which is definably required to enhance the proposed system towards more reliable, robust, and flexible system performance. In order to do this, six systematic measures are described to extend the capability of the proposed framework.

Chapter 2

2. An Introduction to T-S type Fuzzy Model

The representation of decision-making via human reasoning as a way of how human beings process and express the given information can be interpreted as a series of modules in the human brain. As mentioned in Chapter 1, it is a series of these following processes; extracting information by sensing the environment, converting the received information into abstract classes or concepts, linking the processed inputs with the associated reasoning, and performing the action according to the given inputs.

In mathematics, the abstract classes can be represented to sets in set theory. Once the set has been defined, each element of interest becomes either included or excluded from the set. For example, the concept "belongs to", or equivalently "a member of", is the principal mechanism of set theory. In classical set theory, a crisp set is defined as a collection of precise objects and an element in a crisp set either belongs to a set or not. This dichotomization process can be modeled by a *characteristic function* of the crisp set over a certain universe of discourse.

2.1 General Description

In fuzzy set theory [1], a fuzzy set is a collection of distinct elements with a varying "degree of relevance or membership". The *characteristic function* of a fuzzy set, which is known as a *membership function*, takes interval values between 0 and 1, often shown as [0, 1]. The membership values express the degrees with which each object is *compatible* with the properties or features that are distinctive to the collection. In other words, a fuzzy set is a *generalization* of the concept of a set whose characteristic function only takes binary values {0, 1}.

Using the definition and properties of fuzzy sets, the series of human brain behaviors mentioned earlier can be modeled as a process of fuzzy inference. Labeling the categorized classes of abstract sets into the interpreted data sets corresponds to the fuzzification process of inputs using membership functions. Linking the fuzzified inputs with the associated reasoning mechanism can be seen as a step of building the knowledge base in a fuzzy inference system. Processing to take a final action from the associated reasoning can be modeled as performing the fuzzy inference process to produce the final system output.

The fuzzy inference systems as a model of human reasoning have been widely used for a variety of practical industrial areas over the past decades; non-linear system modeling [2], [3], control engineering [5], [6], robotics [8], [9] and so on. There are two major types of fuzzy inference systems; a language-driven type and a data-driven type fuzzy inference system. Language-driven type fuzzy systems, for instance Mamdani type [21], are designed via human language variables and rules, which is based upon the experience of fuzzy experts and/or experts in the specific field. Data-driven type fuzzy inference systems, for example, Takagi-Sugeno (T-S) type [7], are designed based on the experimental input-output data collected from actual experimentation.

In regard to advantages and disadvantages of those two types of fuzzy inference systems, language-driven type fuzzy inference systems are comparatively easier to design and fast to calculate outputs, but their de-fuzzification process is very time-consuming and the systematic fine-tuning is extremely difficult to handle. Meanwhile, data-driven type fuzzy inference systems are excellent in mathematical modeling by the design for their rule consequents, but it is often difficult to assign any appropriate linguistic terms to the rule consequent which is a non-fuzzy membership function as an output variable.

However, in order to utilize the advantage of the mathematical modeling of data-driven type fuzzy inference systems, the Takagi-Sugeno (T-S) type fuzzy model is chosen in this thesis to perform the fuzzy inference process.

In this chapter, the Takagi-Sugeno (T-S) type fuzzy inference model is introduced mentioning its theoretical model, aspects, and analyses as a universal approximator with the associated approximation theories.

2.2 Fuzzy System Models

There are two major models of fuzzy systems; Mamdani [21] and Takagi-Sugeno (T-S) [7] fuzzy systems. The main difference between these two types of fuzzy systems lies in the consequent variable of fuzzy rules. Mamdani type fuzzy systems use linguistic fuzzy sets as consequent variables in fuzzy rules as defined by (2.1), whereas the T-S type fuzzy systems employ a linear combination of input variables as a rule consequent variable as defined by (2.2).

$$R_i : IF\ x_{k1}\ is\ F_{i1}\ AND\ x_{k2}\ is\ F_{i2} \ldots AND\ x_{km}\ is\ F_{im}$$
$$THEN\ y_i\ is\ F^*_i \tag{2.1}$$

$$R_i : IF\ x_{k1}\ is\ F_{i1}\ AND\ x_{k2}\ is\ F_{i2} \ldots AND\ x_{km}\ is\ F_{im}$$
$$THEN\ y_i = c_{i0} + c_{i1}x_{k1} + \ldots + c_{im}x_{km} \tag{2.2}$$

where

R_i *(i=1, 2, …, N)*: the *i*-th fuzzy rule

x_{kj} *(j=1, 2, …, m)*: the *j*-th input feature of the *k*-th pattern vector

F_{ij}: a fuzzy variable of the *j*-th input feature in the *i*-th rule

F^*_i: a linguistic fuzzy set of the *i*-th Mamdani type rule consequent

c_{ij}: a coefficient of inputs of the *i*-th T-S type rule consequent.

Both types of fuzzy models have been deployed widely as effective tools in a variety of practical applications, especially in non-linear system modeling and control system over past decades.

2.2.1 T-S type Fuzzy Model

The T-S type fuzzy model suggested by Takagi and Sugeno [7] is able to represent a general class of non-linear systems. The consequent variables of its fuzzy rules are defined as a linear combination of input variables as defined by (2.3).

$$R_i : IF\ x_{k1}\ is\ F_{i1}\ AND\ x_{k2}\ is\ F_{i2} \ldots AND\ x_{km}\ is\ F_{im}$$
$$THEN\ y_i = c_{i0} + \sum_{j=1}^{m} c_{ij} x_{kj} \tag{2.3}$$

$$y = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}, \quad w_i = \prod_{j=1}^{m} F_{ij}(x_{kj}) \tag{2.4}$$

where

R_i *(i=1, 2, …, N)*: the *i*-th T-S type fuzzy rule

x_{kj} *(j=1, 2, …, m)*: the *j*-th input feature of the *k*-th pattern vector

F_{ij}: a fuzzy variable of the *j*-th input feature in the *i*-th rule

c_{ij}: a coefficient of the T-S type rule consequent

Π: a fuzzy *T-norm* (*'AND'*) operator

w_i: a rule firing strength of the *i*-th rule

y_i: the *i*-th rule output

y: the total output

The T-S type fuzzy model approximates non-linear systems using a combination of several linear systems by decomposing the entire input domain into several partial spaces and representing each input and output space with a linear function. In order to find the coefficients of the linear systems, the least-square fit method has been widely used.

One of the most significant advantages of T-S type fuzzy model is that the representation of the system output is designed using a mathematical equation – a linear combination of inputs, which means it is very effective to describe and calculate the characteristics of non-linear systems. Most of the practical T-S fuzzy systems have used linear functions of input variables as rule consequent variables. The linear rule consequent variable is critical to the practicality and usefulness of T-S fuzzy systems. This is because when non-linear rule consequent variables are used, determining the structures and parameters of the rule consequents properly is extremely difficult. Furthermore, compared with well-established traditional polynomial approximators, the fuzzy system with non-linear rule consequent variables is greatly disadvantageous in terms of computational complexity and practical usefulness. Also using the statistical estimation methods to obtain the coefficients of the consequent variables of the T-S fuzzy rules, the system identification task can be evaluated and enhanced by the statistical analysis towards better system performance.

A collection of input-output data is required to build T-S type fuzzy models due to its mathematical definition of the rule consequent variables. *A prior knowledge* can be gathered by human experts in the corresponding area based on the data observations. Using the collected information, a knowledge base that consists of membership functions and fuzzy rules can be constructed. In general, a supervised or unsupervised clustering method determines the partition of the given knowledge, and membership functions for each feature can be obtained according the resulting partition information. The T-S type fuzzy rules can be obtained as a form of *Generalized Modus Ponen* (GMP) inference through T-S type rule design schemes based on the statistical approaches. Once the T-S type fuzzy inference model is built, when the unknown input is fed into the pre-built T-S type fuzzy system its output is calculated via a method of the generalized de-fuzzifiers. This is a simple description of the systematic mechanism of the T-S type fuzzy inference model.

2.2.2 Design of Membership Functions and Fuzzy Rules

The common types of membership functions are; singletons, triangles, trapezoids, Gaussians, and so on. Every type of membership function has its advantages and disadvantages. For instance, triangular membership function is very easy to implement and fast to calculate on real-time based systems. However, it is very difficult for triangles to adjust adaptively using statistical methods in on/off-line learning schemes towards better system performance due to their discontinuity in their mathematical form. In the context of control systems, obviously the priority of the system often lies on the speed of real-time system performance in most of the industrial control systems. Thus the triangle type membership function has been employed widely in control engineering.

For Gaussian functions, it takes time to calculate their output, but they have an advantage for describing the gathered data as a naturally distributed statistical model, and also the exponential term in Gaussian functions allows the adaptive mechanism to adjust them with the statistical learning functions. This property of Gaussian basis functions provides higher accuracy and more flexibility to model the non-linear systems in system modeling.

To design the membership functions for input data, an unsupervised clustering method can be utilized for practical systems. The Fuzzy C-Means (FCM) clustering approach [22], for example, has been used in many papers [23], [24], [25] to find the *C* number of adaptive fuzzy clusters for each feature of input data. Once the fuzzy clusters are obtained, the type of membership functions can be chosen to model membership functions for fitting the processed membership values.

In regard to the design of fuzzy rules, the antecedent part of the T-S type fuzzy rules is composed with linguistic fuzzy sets, whereas the consequent part of them is defined as the linear combination of input variables as defined in (2.3). The antecedent variables of the T-S type fuzzy rules can be designed using a particular type of membership function and a fitting process. However, it should be noted that a particular type of the fuzzy membership function for antecedent parts should be determined according to the system objectives and the characteristics of the T-S type fuzzy system towards a specific goal. The consequent part, as stated earlier, is defined as a weighted summation of inputs to represent the non-linear characteristic functions as a general class. In order to estimate the coefficients of the rule consequent, the least squares estimation has been widely used.

2.2.3 Fuzzy Inference Process for the T-S type Fuzzy Model

Using the designed membership functions and fuzzy rules, the fuzzy inference process in the T-S type model can approximate the unknown input data. When the unknown input is fed into the T-S type fuzzy inference system, each feature value of the unknown input vector is fuzzified, i.e., converted to a fuzzy number, through their membership functions in the knowledge base. The fuzzified inputs are calculated with the '*AND*' operator using a fuzzy T-*norm* operator in which the algebraic minimum function is generally employed. Its output is then linked with '*THEN*' operator, which is an implication operator to calculate the level of the rule firing strength for each rule. The rule output for each rule now can be determined by its linear combination equation as defined (2.3). A total output is obtained by (2.4) as a special case of the 'generalized de-fuzzifiers'. This fuzzy inference process for the T-S type fuzzy model is shown in Figure 2.1.

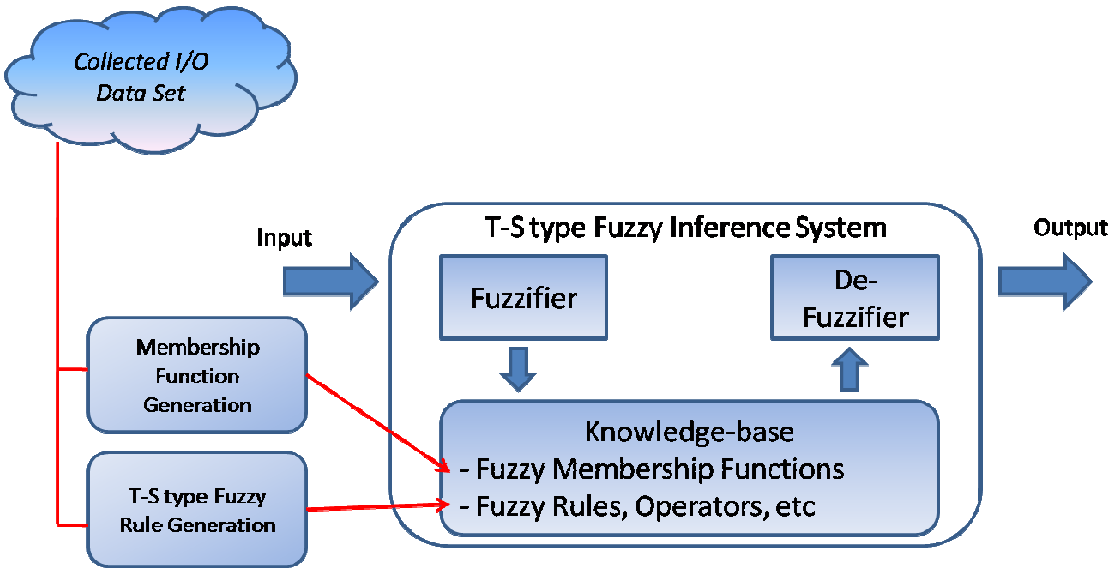


Figure 2.1 Fuzzy inference process for the T-S type fuzzy model

2.3 General T-S type fuzzy models are Universal Approximators

The issue of fuzzy systems as universal approximators has been addressed as one of the important research interests in the past decade [26], [27], [28], [29], [30], [31]. As Ying mentioned in [30], the basic question is "Can fuzzy systems approximate any real continuous functions to any degree of accuracy on a compact domain?" It is obvious that the answer of this question is crucial to theoretical and practical aspects of fuzzy systems. This approximation question asks in the context of fuzzy control whether or not a fuzzy controller can be constructed to approximate any continuous non-linear control solution. In the context of modeling, the question of interest is whether or not a fuzzy model is capable of approximating any physical dynamic model that is continuous and non-linear.

In regard to T-S type fuzzy models as universal approximators, the questions include;

1) Are T-S fuzzy systems with linear rule consequent universal approximators?
2) What are the sufficient and necessary conditions for T-S fuzzy systems with linear rule consequent as universal approximators?

One of the best approaches to answer these questions has been done in Ying's study in [29], [30], [31] proving that the general class of T-S type fuzzy systems can uniformly approximate 1) any polynomial arbitrarily well and 2) any continuous function with arbitrarily high precision for both Single-Input-Single-Output (SISO) and Multi-Input-Single-Output (MISO) type T-S fuzzy systems by utilizing the Weierstrass approximation theorem [32]. In this section, the proof and the associated theorems related to the questions mentioned above are reviewed according to Ying's investigation.

2.3.1 Configuration for General T-S type Fuzzy Systems with Linear Rule Consequent

The general T-S type fuzzy systems in Ying's investigation uses a *p*-dimensional continuous-time or discrete-time multi input vector $x(t)$ as defined by (2.5).

$$x(t) \equiv (x_1(t), x_2(t), \cdots, x_p(t)) \tag{2.5}$$

where

t: time and $-1 \leq x_l(t) \leq 1$ ($l = 1, 2, \ldots, p$)

For the fuzzification process, $N = 2n+1$ ($n \geq 1$) number of fuzzy sets, denoted F_j, are used for each input with arbitrary continuous membership function type. The $2n$ equal intervals are partitioned in [-1, 1] for each input, each of which is $[j/n, (j+1)/n]$ ($j = 0, \pm 1, \ldots, \pm n$) so that the $(2n + 1)^p = N^p$ fuzzy rules are used to cover all F_{jl}.

2.3.2 General MISO T-S Fuzzy Systems

The N^p numbers of MISO T-S type fuzzy rules are expressed with linear rule consequents as shown in (2.6).

$$\begin{aligned} R_i: \quad & IF\ x_1(t)\ is\ F^i{}_{j1}\ AND\ x_2(t)\ is\ F^i{}_{j2}\ AND \cdots AND\ x_p(t)\ is\ F^i{}_{jp} \\ & THEN\ y_i = c_{i0} + c_{i1}x_1(t) + c_{i2}x_2(t) + \cdots + c_{ip}x_p(t) \end{aligned} \tag{2.6}$$

where

R_i $(i=1, 2, \ldots, N^p)$: the *i*-th MISO T-S type fuzzy rule

F^i_{jl} ($j = 0, \pm 1, \ldots, \pm n$, $l = 1, 2, \ldots, p$): the *j*-th fuzzy set for the *l*-th input vector in the *i*-th rule

y_i: the *i*-th rule output

c_{il}: the design parameters whose values are determined by the fuzzy system developer.

There are p+1 design parameters in each rule. Thus, there are in total $(p + 1)N^p$ system parameters for N^p rules. Obviously in different rules, the values of these parameters are different.

In order to reduce the number of design parameters, Ying has proposed a simplified linear T-S rule consequent in [33], [34] using SISO T-S type model as an example, which is defined by (2.7).

$$R_i : \quad IF\ x(t)\ is\ F_{ji} \quad THEN\ y_i = k_i\,(a + bx(t)) \tag{2.7}$$

where

R_i: the i-th T-S type fuzzy rule (i = 1, 2, …, M)

y_i: the i-th rule output

a, b, k_i : design parameters whose values are determined by the fuzzy system developer.

As mentioned in [30], the major advantage in using the simplified T-S rule consequent over the original one is the significant reduction in the number of design parameters. Because all the rule consequents in (2.7) used the same linear function $a + bx$ and all the rules are proportional to each other. The reduction to the example SISO model in (2.7) is by a factor of $(M - 2) / 2M$, which is almost 50% for larger M. In order to illustrate another extended example of the reduction, a SISO fuzzy system that uses ten fuzzy rules is considered (M=10) as an example. The original T-S rule consequent will require 20 parameters whereas the simplified consequent only 12, which means a 40% reduction.

For MISO type T-S fuzzy systems, the reduction is even greater; the more the number of input variables, the greater the reduction. Applying this simplified model for reduction to the MISO T-S type model (2.7), the reduction is by a factor of $\{ pN^p – (p +1) \} / \{(p +1) N^p \}$. For instance, if the input vector had nine multi-variables, or p = 9, then the parameters of simplified MISO T-S type fuzzy rule consequent have almost a 90% reduction. This kind of parameter reduction is not only desirable, but also necessary in many practical applications, especially control applications.

Also the simplified linear T-S rule consequent is a special case of the original linear T-S type rule consequent [30] as shown in (2.8).

$$
\begin{aligned}
a_i &= k_i a \\
b_i &= k_i b
\end{aligned} \tag{2.8}
$$

In regard to the defuzzification process, using the general rule consequent (2.6), the generalized defuzzifier [35] is used to calculate the total system output *y*, which actually is a mapping ψ_n: [-1, 1] → (-∞, +∞), as defined by (2.9).

$$
\psi_n(x) \equiv y = \frac{\sum_{i=1}^{(2n+1)^p} (\mu_i)^\alpha \cdot (c_{i0} + c_{i1}x_1 + c_{i2}x_2 + \cdots + c_{ip}x_p)}{\sum_{i=1}^{(2n+1)^p} (\mu_i)^\alpha} \tag{2.9}
$$

where

μ_i : the membership for the rule consequent in *i*-th rule.

$\mu_i = \prod \mu_{F\,jl}^{\,i}$, Π: a fuzzy *T-norm* operator (‘AND’ fuzzy logic)

Different defuzzification results can be obtained by using different α values where $0 \le \alpha \le +\infty$. The most widely used centroid defuzzifier is a special case of this generalized defuzzifier when $\alpha = 1$, and the popular mean of maximum defuzzifier is another special case when $\alpha = \infty$. The $\psi_n(x)$ is a function sequence with respect to *n*. The mapping $\psi_n(x)$ will be used to represent the general MISO T-S type fuzzy systems.

2.3.3 General T-S Fuzzy Systems are Universal Approximators

In order to prove that the general T-S fuzzy systems with T-S linear consequent are universal approximators, firstly Ying has proved that the general MISO T-S fuzzy models can uniformly approximate any polynomial to any degree of accuracy. Then utilizing the Weierstrass approximation theorem [32], he has proved that the general T-S fuzzy models can uniformly approximate any multivariate continuous function with arbitrary precision. In this section, the proved theorems and associated approximation theorems will be reviewed. For further details of proofs, refer to the literature [29], [30], [31].

Theorem 1

$\psi_n(x)$ can uniformly approximate, with arbitrarily high precision, any polynomial $P_h(x)$ defined on [-1, 1].

$$P_h(x) = \sum_{i=0}^{h} \beta_i x^i \tag{2.10}$$

where

h: the order of the polynomial.

The final formula of the proof for this theorem is as follows.

$$n^* > \frac{|\beta_1| + \sum_{i=1}^{h} |\beta_i| \, (2^i - 1)}{\varepsilon}, \quad \forall n > n^* \tag{2.11}$$

where

ε: a positive approximation error bound

n^* : a positive integer based on a given function and approximation error ε

In other words, proving Theorem 1 is equivalent to proving that there exists a positive integer n^* that satisfies the equation (2.11). By having derived (2.11), it has been proven that the general T-S type fuzzy systems can approximate any polynomials uniformly with arbitrarily high accuracy. Now the Weierstrass approximation theorem [32] is briefly mentioned as a basis to the next theorem.

Weierstrass Approximation Theorem

To any continuous function G(x) on a closed interval, given approximation error bound $\varepsilon>0$, there always exists a polynomial that can approximate G(x) uniformly with the desired accuracy. In general, the smaller the ε, the higher the polynomial degree.

Theorem 2

The general MISO T-S type fuzzy systems with linear rule consequent can uniformly approximate any continuous function on a closed interval to any degree of accuracy.

The proof for theorem 2 is as follows. The polynomial P_h can uniformly approximate *G(x)* with arbitrary accuracy according to the Weierstrass approximation theorem. The result of the proof is as follows in (2.12), which means $\psi_n(x)$ can uniformly approximate *G(x)* arbitrarily well.

$$\|\psi_n - G\| \le \|\psi_n - P_h\| + \|P_h - G\| = \varepsilon, \quad \varepsilon = \varepsilon_1 + \varepsilon_2, \forall \varepsilon_1 > 0, \forall \varepsilon_2 > 0 \tag{2.12}$$

It has been proved that the general MISO T-S fuzzy models are universal approximators.

Chapter 3

3. Rough Sets Theory

Knowledge discovery is a process that combines techniques from machine learning, statistics, pattern recognition, fuzzy and rough sets, etc. to extract knowledge, or information, from vast amounts of data. Often it is used to support human decision-making processes or to explain observed phenomena. Knowledge discovery is a process that helps to make sense of data in more readable and applicable form. The knowledge discovery process and its data mining tools are becoming the focus of many fields, particularly in data-rich and knowledge-poor processing scenarios.

This kind of process usually starts with sampling, feature selection or discretization, transformation or projection, dimensionality reduction, extraction of data, and physical phenomena models, usually followed by algorithms. These algorithms here generate hypotheses about extracted data. These hypotheses are used as new extracted knowledge. Sufficient methods of extracting knowledge from database or multivariate experimental data sets belong to basic information processing steps. In particular, consideration of implicit, imprecise, and insufficient knowledge in databases or experimental data sets is of importance in developing knowledge-based systems. Another fundamental issue in multi-dimensional pattern processing is feature extraction and reduction relevant for robust prediction and performance.

3.1 Introduction

The rough sets theory has been developed for knowledge discovery in databases and experimental data sets. This theory provides a powerful foundation to reveal and discover important structures in data and to classify complex objects. The attribute-oriented rough sets technique reduces the computational complexity of learning processes and eliminates the unimportant or irrelevant attributes so that the knowledge discovery in the database or in experimental data sets can be learned efficiently.

Rough Set theory was introduced by Zdzislaw Pawlak [36] to provide a systematic framework for studying imprecise and insufficient knowledge. The information system proposed by Pawlak is for representing knowledge and discovering relationships in data. Rough set theory has been studied in medical databases analysis [40], [41], [42], image analysis for medical applications [43], [44], [45], decision support systems [46], [47], pattern recognition [48], [49], [50], and machine learning [51], [52] and so on. Using rough sets has been shown to be very effective for revealing relationships within imprecise data, discovering dependencies among objects and attributes, evaluating the classificatory importance of attributes, removing data redundancies and thus reducing the size of information systems, and generating decision rules.

Some classes or categories of objects in an information system cannot be distinguished in terms of the available attributes. They can only be roughly or approximately defined. The idea of rough sets is based on equivalence relations which partition a data set into equivalence classes, and consist of the approximations of a set by a pair of sets, called lower and upper approximations. The lower approximation of a set of object (a concept) contains all objects that, based on the knowledge of a given set of attributes, can be classified as certainly belonging to the concept. The upper approximation of a set contains all objects that cannot be classified categorically as not belonging to the concept. A rough set is defined as an approximation of a set using a pair of sets: the upper and lower approximations of the set.

The rough sets theory also deals with information that can be represented in a form of a table. This table consists of objects (or cases) and attributes. The entries in the table are the categorical values of the features (attributes) and, for some information systems, possibly also associated classes (categories). Many data processing problems can be easily converted into a data table representation and analysis. By processing information using the rough sets theory, a classification of objects in an information system can be discovered.

The rough sets theory can be applied to a variety of information processing. For example, it can be used for the followings.

- Creating a decision table representing an information system containing uncertain or imprecise data
- Analyzing the relationships of data in given knowledge
- Computing lower and upper approximations of sets
- Evaluating dependencies of attributes in data sets

- Computing a quality and an accuracy of approximation
- Calculating reducts as sets of minimal number of attributes describing concepts
- Reducing data with information preservation by removing superfluous attributes
- Reasoning with uncertainties
- Deriving decision algorithms as a set of decision rules.

The basic definition and the most of the all functionalities of rough set theory are reviewed in this chapter in order to guide the reader easily by following the example-based review.

3.2 Information System

An *information system* is a representation of data gathered from measurements of some physical phenomena, for example, speech signals, sequences of images, industrial process signals, and so on. An information system denoted by *S* is composed of four elements as defined in (3.1).

$$S = < U, Q, V, f > \tag{3.1}$$

where

U: a *closed universe* which is a finite non-empty set of N objects $\{x_1, x_2, \ldots, x_N\}$

Q: a finite and non-empty set of *n attributes* $\{q_1, q_2, \ldots, q_n\}$

$V = \bigcup_{q \in Q} Vq$: a *domain of attributes*

$f: U \times Q \rightarrow V$: a total *decision function* which maps elements of attributes

Any pair (q, v) for $q \in Q$, $v \in Vq$ is called the *descriptor* in an information system *S*. The information system can be represented as a finite data table, in which the columns are labeled by *attributes*, the rows by *objects* and the *entry* in column *q* and row *x* has the *value f(x, q)*. Each row in the table describes the information about some object in *S*. Any non-empty set of objects *X* is called a concept in *S*. A concept might have a certain meaning. For example, as shown in Example 3.1, in a medical data set with tests and diagnoses, one can define a concept as a set of objects representing sick patients.

[Example 3.1]

Suppose a simple example of a data set called MEDICAL representing medical findings.

Table 3.1 A medical data set MEDICAL (modified from [53])

Object	*Attributes*			
U	*c1*	*c2*	*c3*	*d*
x_1	*0*	*L*	*N*	*0*
x_2	*0*	*H*	*N*	*1*
x_3	*0*	*H*	*N*	*1*
x_4	*0*	*L*	*N*	*0*
x_5	*1*	*H*	*Y*	*1*
x_6	*1*	*H*	*Y*	*1*
x_7	*1*	*H*	*Y*	*1*
x_8	*2*	*L*	*Y*	*0*
x_9	*2*	*L*	*Y*	*1*
x_{10}	*2*	*H*	*Y*	*0*
x_{11}	*2*	*H*	*Y*	*1*

In the information system S describing this data set, the universe U consists of ten objects, $U = \{x_1, x_2 \ldots x_{10}\}$ each representing one patient. Each patient is described by the set of three attributes $Q = \{q_1, q_2, q_3, q_4\} = \{c_1, c_2, c_3, d\}$, with discrete values (numerical and symbolic), representing results of the medical tests and diagnoses. The set of all discrete (numerical) values of the attribute *c1* is $V_{c1} = \{0, 1, 2\}$. The second attribute c_2 takes two discrete non-numerical values $V_{c2} = \{L, H\}$ (L=*LOW*, H=*HIGH*). The third attribute c_3 is with two discrete non-numerical values $V_{c3}= \{N, Y\}$ (N=*NO*, Y=*YES*). The fourth attribute d, with two binary values $V_d = \{0, 1\}$, represents an expert's (doctor's) decision, being a diagnosis about a certain disease based on test results. The decision attribute $d = 0$ denotes the diagnosis that a patient does not have a disease, and $d = 1$ that he or she does. Values of information function $f(x, q)$ are included in Table3.1. For example, for the object x_1 and the attribute c_1, the information function values if $f(x_1, c_1) = 0$. A set of objects $\{x_2, x_3, x_5, x_6, x_7, x_9, x_{10}\}$ can be defined as an example of a concept (sick patients) in the considered information system.

3.3 Indiscernibility Relation

Let $S =< U, Q, V, f >$ be an information system, $A \subseteq Q$ be a subset of attributes, and $x, y \in U$ be objects. Then objects x and y are *indiscernible* by the set of attributes A in S

(denoted by $x\tilde{A}y$) iff *f*(*x*, *a*) = *f*(*y*, *a*) for every $a \in A$. For any given subset of attributes $A \subseteq Q$, the *IND* (*A*), denoted by $\tilde{A}$, is an *equivalence relation* on universe *U* and is called an *indiscernibility relation*. The indiscernibility relation, *IND* (*A*) is defined as follows in (3.2).

$$IND(A) = \{(x, y) \in U \times U : \forall a \in A, f(x,a) = f(y,a)\} \tag{3.2}$$

If the pair of objects (*x*, *y*) belongs to the relation $IND(A) = ((x, y) \in IND(A)))$ then objects *x* and *y* are called *indiscernible* with respect to *A*. In other words, one cannot distinguish object *x* from *y* in terms of attributes from set *A* only. The indiscernibility relation *IND (A)* as a binary equivalence relation splits the given universe *U* into a family of equivalence classes $\{X_1, X_2, \ldots, X_r\}$. The family of all equivalence classes $\{X_1, X_2, \ldots, X_r\}$, defined by the relation *IND* (*A*) on *U*, generates a partition of *U* and it is denoted by A^*, The family of equivalence classes A^* is also referred as classification and also denoted by *U / IND* (*A*).

Objects belonging to the same equivalence class X_i are indiscernible; otherwise objects are discernible with respect to the attribute subset *A*. The equivalence classes X_i, $i=1,2,..,r$ of the relation *IND (A)* are called *A-elementary sets* in an information system *S*. An *A-elementary set* $[x]_A$, or an equivalence class, including an object *x* is defined by (3.3).

$$[x]_A = \{y \in U : xIND(A)y, or, x\tilde{A}y\} \tag{3.3}$$

For a given information system *S*, a given subset of attributes $A \subseteq Q$ generates an indiscernibility relation *IND* (*A*) (an equivalence relation). An ordered pair *AS* = (*U*, *IND* (*A*)) is called an *approximation space*. Any finite union of elementary sets in *AS* is called a *definable set* of a *composed set* in *AS*. $Des_A(X)$ denotes the description of A-elementary set $X \in A^*$ (an equivalence class) and it is defined as follows in equations (3.4) and (3.5). Equation (3.4) is often denoted as equation (3.5).

$$Des_A(X) = \{(a,b) : f(x,a) = b, \forall x \in X, a \in A\} \tag{3.4}$$

$$Des_A(X) = \{(a = b) : f(x,a) = b, \forall x \in X, a \in A\}. \tag{3.5}$$

[Example 3.2]

Let us analyze an information system, MEDICAL from Table 3.1 and assume only results of tests are considered, representing by the subset of attributes $A = \{c_1, c_2, c_3\}$ and contained in the reduced Table 3.2.

Table 3.2 The MEDICAL data set with the reduced attribute set $A = \{c_1, c_2, c_3\}$

Object	*Attributes*		
U	*c1*	*c2*	*c3*
x_1	*0*	*L*	*N*
x_2	*0*	*H*	*N*
x_3	*0*	*H*	*N*
x_4	*0*	*L*	*N*
x_5	*1*	*H*	*Y*
x_6	*1*	*H*	*Y*
x_7	*1*	*H*	*Y*
x_8	*2*	*L*	*Y*
x_9	*2*	*L*	*Y*
x_{10}	*2*	*H*	*Y*
x_{11}	*2*	*H*	*Y*

Equivalence classes:

$E_1 = [x_1]_A = [x_4]_A = \{x_1, x_4\}$

$E_2 = [x_2]_A = [x_3]_A = \{x_2, x_3\}$

$E_3 = [x_5]_A = [x_6]_A = [x_7]_A = \{x_5, x_6, x_7\}$

$E_4 = [x_8]_A = [x_9]_A = \{x_8, x_9\}$

$E_5 = [x_{10}]_A = [x_{11}]_A = \{x_{10}, x_{11}\}$

From Table 3.2 it can be seen that objects can be divided into five disjoint groups according to equal values of attributes c_1, c_2 and c_3 from the subset A. Objects in the same group have the same values for all attributes as the other objects from this group. For example, in the first group it has two objects x_1, x_4 since no other objects have values $c_1 = 0$, $c_2 = L$ and $c_3 = N$ for attributes from A. The object x_1 belongs to the equivalence class $E_1 = [x_1]_A = [x_4]_A = \{x_1, x_4\}$. Objects x_2 and x_3 with equal values for all attributes $c_1 = 0$, $c_2 = H$, $c_3=N$ form the second group. It can be observed that objects in this group cannot be distinguished based on attributes c_1, c_2 and c_3 from the set A only. They belong to the equivalence class $E_2 = [x_2]_A = [x_3]_A = \{x_2, x_3\}$. Similarly, it can be found that other equivalence classes in S for set A; $E_3 = [x_5]_A = [x_6]_A = [x_7]_A = \{x_5, x_6, x_7\}$, $E_4 = [x_8]_A = [x_9]_A = \{x_8, x_9\}$, $E_5 = [x_{10}]_A = [x_{11}]_A = \{x_{10}, x_{11}\}$.

As shown, a subset of attributes $A \subseteq Q$ imposes an indiscernibility relation *IND* (*A*) on the whole set of objects from the universe *U*. It can be implied that a relation *IND* (*A*) as; All pairs of objects (x_i, x_j) in *S* for which values of all attributes from *A* are all equal.

3.4 Discernibility Matrix

Frequently discernibility of objects is more interesting than specific values of attributes. In these cases an information system may be represented as a discernibility matrix. Skowron and Rauszer have introduced two notions [37], the discernibility matrix and the discernibility function, which help to construct efficient algorithms related to the generation of minimal subsets of attributes sufficient to describe all concepts in a given information system. With these two notions, the differences between the attributes of each pair of objects can be stored into a matrix called a *discernibility matrix*. The discernibility matrix contains fewer data than those of an information system but holds all essential information used to check whether a set of attributes is a minimal one that describes concepts.

Let $S = < U, Q, V, f >$ be an information system and assume that $U = \{x_1, x_2, \ldots, x_N\}$, $Q = \{a_1, a_2, \ldots, a_n\}$. A discernibility matrix *M(Q)* for an information system *S* with the set of attributes *Q* is a square $N \times N$ dimensional matrix, with rows and columns labeled by objects x_i ($i=1,2,\ldots N$). Each entry m_{ij} of a discernibility matrix (for a given row *i* and a column *j* representing two objects x_i and x_j from *U*) is a subset of attributes which discerns these objects. Therefore, a discernibility matrix can be defined by (3.6).

$$m_{ij} = \begin{cases} 0 & x_i, x_j \in \textit{the same equivalence class of IND(Q)} \\ \{a \in Q : f(x_i, a) \neq f(x_j, a)\} & x_i, x_j \in \textit{different equivalence classes of IND(Q)} \end{cases} \quad (3.6)$$

where

$$x_i, x_j \in U$$

The entry m_{ij} contains all these attributes whose values are not identical for both x_i and x_j, which means that x_i, x_j belong to different classes of partition generated by *IND* (*Q*). The discernibility matrix *M*(*Q*) is symmetric and $m_{ii} = 0$, thus it is sufficient to compute only entries in the lower triangle of *M* (*Q*), i.e., the m_{ij} with $0 \leq j < i \leq N-1$.

A discernibility function f_S for an information system S is a Boolean function of n Boolean variables $\overline{a_1}, \overline{a_2}, \ldots, \overline{a_n}$ corresponding to the attributes $a_1, a_2, \ldots, a_n$ respecttively. It is defined by equation (3.7).

$$f_A(\overline{a_1}, \overline{a_2}, \ldots, \overline{a_n}) = \wedge\{\vee(m_{ij}) : 1 \leq j < i \leq n, m_{ij} \neq \phi\} \quad (3.7)$$

where

$\vee(m_{ij})$: A disjunction of all variables $\overline{a}$ such that $a \in m_{ij}$

[Example 3.3]

Suppose a given information system MEDICAL as shown in Table 3.2. The discernibility matrix M (Q) can be obtained as shown in Table 3.3. ($m_{ii} \neq 0$, $m_{ij} = m_{ji}$ for i, j=1,…,6)

Table 3.3 Discernibility Matrix *M(Q)*

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{11}	x_{11}
x_1											
x_2	c_2	Ø									
x_3	c_2	Ø									
x_4	Ø	c_2	c_2								
x_5	$c_1c_2c_3$	c_1c_3	c_1c_3	$c_1c_2c_3$							
x_6	$c_1c_2c_3$	c_1c_3	c_1c_3	$c_1c_2c_3$	Ø						
x_7	$c_1c_2c_3$	c_1c_3	c_1c_3	$c_1c_2c_3$	Ø	Ø					
x_8	c_1c_3	$c_1c_2c_3$	$c_1c_2c_3$	c_1c_3	c_1c_2	c_1c_2	c_1c_2				
x_9	c_1c_3	$c_1c_2c_3$	$c_1c_2c_3$	c_1c_3	c_1c_2	c_1c_2	c_1c_2	Ø			
x_{10}	$c_1c_2c_3$	c_1c_3	c_1c_3	$c_1c_2c_3$	c_1	c_1	c_1	c_2	c_2		
x_{11}	$c_1c_2c_3$	c_1c_3	c_1c_3	$c_1c_2c_3$	c_1	c_1	c_1	c_2	c_2	Ø	

The discernibility function is as follows.

$$f_S(c_1, c_2, c_3) = c_1 \wedge c_2 \wedge (c_1 \vee c_2) \wedge (c_1 \vee c_3) \wedge (c_1 \vee c_2 \vee c_3)$$

3.5 Decision Tables

Information systems can be designed as a decision table if the attribute set Q is divided into two disjoint sets which are condition attribute set C and decision attribute set D, i.e.

$Q = C \cup D$. For example, for a data set gathered for a classification task, a set C of condition attributes may represent elements of a pattern x describing an object and a set D may represent a classification decision, for instance, a categorical class assigned to an object. In a given information system S, a decision table DT is defined by (3.8).

$$DT = < U, C \cup D, V, f > \quad (3.8)$$

where

U: a *closed universe* which is a finite non-empty set of N objects$\{x_1, x_2, \ldots, x_N\}$

C: a non-empty set of condition attributes (features of input pattern vectors)

D: a non-empty set of decision attributes (target classes)

V: a domain of attributes

f: a total decision function, or an information function in a DT

A decision table is called *deterministic* if each object's decision attributes values are uniquely specified by a particular object's condition attributes. If a number of decision attribute values may be taken for a given condition attribute, it is called *non-deterministic*. Some of non-deterministic decision tables may be decomposed into two sub-tables; deterministic and *totally non-deterministic*. A totally non-deterministic decision table does not contain a deterministic sub-table.

[Example 3.4]

The MEDICAL data set from Example 3.1 can be interpreted as a decision table as shown in Table 3.4.

Table 3.4 A decision table MEDICAL

Object	*Attributes*			
	C (medical diagnoses)			*D (disease class)*
U	*c1*	*c2*	*c3*	*d*
x_1	*0*	*L*	*N*	*0*
x_2	*0*	*H*	*N*	*1*
x_3	*0*	*H*	*N*	*1*
x_4	*0*	*L*	*N*	*0*
x_5	*1*	*H*	*Y*	*1*
x_6	*1*	*H*	*Y*	*1*
x_7	*1*	*H*	*Y*	*1*
x_8	*2*	*L*	*Y*	*0*
x_9	*2*	*L*	*Y*	*1*
x_{10}	*2*	*H*	*Y*	*0*
x_{11}	*2*	*H*	*Y*	*1*

3.6 Approximation of Sets

Some subsets of objects in an information system cannot be distinguished in terms of the available attributes. They can only be approximately defined. The idea of *rough sets* consists of an approximation of a set by a pair of sets, called a *lower* and an *upper approximation* of this set.

A given subset of attributes $A \subseteq Q$ in a given information system *S*, determines the approximations space *AS* = (*U*, *IND* (*A*)) in *S*. For a given $A \subseteq Q$ and $X \subseteq U$, an *A-lower approximation* $\underline{A}X$ and an *A-upper approximation* $\overline{A}X$ of set *X* in *AS* are defined as follows in (3.9) and (3.10).

$$\underline{A}X = \{x \in U : [x]_A \subseteq X\} = \bigcup \{Y \in A^* : Y \subseteq X\} \tag{3.9}$$

$$\overline{A}X = \{x \in U : [x]_A \cap X \neq \phi\} = \bigcup \{Y \in A^* : Y \cap X \neq \phi\} \tag{3.10}$$

where

$[x]_A$: an equivalence class which contains *x* on an equivalence relation *IND* (*A*)

A lower approximation $\underline{A}X$ of a set *X* is a union of all equivalence classes that are subsets of *X*. For any $x \in \underline{A}X$, it is certain that *x* belongs to *X*. In other words, a lower approximation $\underline{A}X$ of a set *X* contains all objects that, based on the knowledge of attributes *A*, can be classified as certainly belonging to the concept *X*.

An upper approximation $\overline{A}X$ of a set *X* is a union of all equivalence classes that have non-empty intersections with *X*. For any $x \in \overline{A}X$, it can be said that *x* can possibly belongs to *X*. In other words, an upper approximation $\overline{A}X$ of a set *X* contains all objects that cannot be classified as not belonging to the concept *X*.

An *A-boundary* region of a set *X* in *AS*, as a *doubtful region* of *IND* (*A*) is defined as follows in (3.11). For any $x \in U$ belonging to $BN_A(X)$, it is impossible to determine that *x* belongs to *X* based on the description of elementary sets of *IND* (*A*).

$$BN_A(X) = \overline{A}X - \underline{A}X \tag{3.11}$$

An *A-lower approximation* of a set *X* is a *possibly* (the greatest) definable set in *A* of a set *X* and an *A-upper approximation* of a set *X* is a *certainly* (the smallest) definable set in *A* of a set *X*. An *A-boundary* is a doubtful region in *A* of a set *X*.

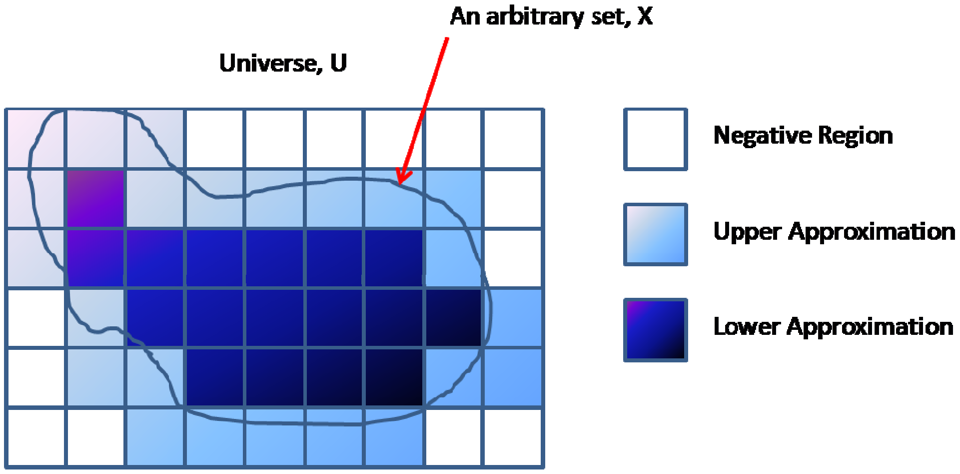


Figure 3.1 A set approximation of an arbitrarily set *X* in *U*

Given an approximation space *AS* for $A \subseteq Q$ and a set $X \subseteq U$, the universe can be partitioned into the following three regions as follows:

1. An *A-positive* region $POS_A(X)$ of *X* in *S*: $\underline{A}X$ (3.12)
2. An A-*boundary* region $BN_A(X)$ of *X* in *S*: $BN_A(X) = \overline{A}X - \underline{A}X$ (3.13)
3. An A-*negative* region $NEG_A(X)$ of *X* in *S*: U - $\overline{A}X$ (3.14)

If $\underline{A}X = \overline{A}X$ then it can be said that $X \subseteq U$ is *A-exactly* approximated in *AS*. In this case the *A-boundary* region $BN_A(X)$ = 0. If $\underline{A}X \neq \overline{A}X$ then $X \subseteq U$ is *A-roughly* approximated in *AS* and the *A-boundary* region $BN_A(X) \neq 0$. The *A-boundary* of *A-exact* set is an empty set.

Here are some properties about the lower and the upper approximation of a set X in A.

1. $\underline{A}X \subseteq X \subseteq \overline{A}X$
2. $\underline{A}\phi = \overline{A}\phi = \phi$
3. $\underline{A}U = \overline{AU} = U$
4. $\underline{A}(X \cup Y) \supseteq \underline{A}X \cup \underline{A}Y$
5. $\overline{A}(X \cup Y) = \overline{A}X \cup \overline{A}Y$

6. $\underline{A}(X \cap Y) = \underline{A}X \cap \underline{A}Y$

7. $\overline{A}(X \cap Y) \subseteq \overline{A}X \cap \overline{A}Y$

8. $\underline{A}(-X) = -\overline{A}X$

9. $\overline{A}(-X) = -\underline{A}X$

10. $\underline{A}\underline{A}X = \overline{A}\underline{A}X = \underline{A}X$

11. $\overline{A}\overline{A}X = \underline{A}\overline{A}X = \overline{A}X$

[Example 3.5]

Suppose a subset X_1 of objects from U in an information system S from Table MEDICAL, representing sick patients (a "*sick*" concept: d = 1).

$X_1 = \{x_2, x_3, x_5, x_6, x_7, x_9, x_{11}\}$

According to the definition of the lower and the upper approximation of a set X_1, based on a subset of attributes $A = \{c_1, c_2, c_3\}$, the lower approximation is the largest composed set of *A-elementary* sets in S that is contained in the subset X_1.

$\underline{A}X_1 = \{\{x_2, x_3\} \cup \{x_5, x_6, x_7\}\} = \{x_2, x_3, x_5, x_6, x_7\}$

The lower approximation contains all A-elementary sets such that every element of the elementary set is also an element of X_1. A lower approximation consists of patients that surely have a disease.

The upper approximation of set X_1 is the smallest composed set of A-elementary sets in S that contain a subset X_1. An upper approximation consists of patients that possibly have a disease.

$\overline{A}X_1 = \{\{x_2, x_3\} \cup \{x_5, x_6, x_7\} \cup \{x_8, x_9\} \cup \{x_{10}, x_{11}\}\} = \{x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}$

The *A-boundary* region (*A-doubtful* region of *IND* (A)) of the set X_1 in S based on A, is

$BN_A(X_1) = \overline{A}X_1 - \underline{A}X_1 = \{x_8, x_9, x_{10}, x_{11}\}$.

This boundary region consists of the composed set of A-elementary sets from S whose elements, based on the subset of attributes A, cannot be classified as belonging to X_1 or not.

In rough sets theory, a set X is either definable or un-definable. A set $X \subseteq U$ is definable in A, denoted by *A-definable,* iff $\underline{A}X = \overline{A}X$, otherwise X is not definable, denoted by *A-non-definable*. In other words, a set X is definable if every object $x \in U$

can be determined with certainty whether $x \in X$ or not. Then the lower approximation of X will be equal to the upper approximation of X, and the boundary of X will be equal to the empty set.

1. A set X is ***roughly*** *A-definable* iff $\underline{A}X \neq \phi$ and $\overline{A}X \neq U$.

   The lower and upper approximation of a set X can be defined. Thus it is possible to decide for some elements of U whether they belong to X or $-X$.

2. A set X is ***externally*** *A-non-definable* in S iff $\underline{A}X \neq \phi$ and $\overline{A}X = U$.

   It cannot be said that any $x \in U$ is not an element of X. Thus it can be determined that for some elements of U they belong to X, but it cannot be said that for any element of U they belong to $-X$ or not.

3. A set X is ***internally*** *A-non-definable* in S iff $\underline{A}X = \phi$ and $\overline{A}X \neq U$.

   It cannot be said that any $x \in U$ is an element of X. Thus it can be determined that for some elements of U they belong to $-X$, but it cannot be said that for any element of U they belong to X or not.

4. A set X is ***totally*** *A-non-definable* in S iff $\underline{A}X = \phi$ and $\overline{A}X = U$.

   The approximations cannot be defined at all. For any element $x \in U$, it cannot be decided to belong to X or $-X$.

[Example 3.6]

From example 3.2, the equivalence classes of MEDICAL data set with $A = \{c_1, c_2, c_3\}$ are as follows.

$E_1 = \{x_1, x_4\}$ $\quad$ $E_2 = \{x_2, x_3\}$

$E_3 = \{x_5, x_6, x_7\}$ $\quad$ $E_4 = \{x_8, x_9\}$

$E_5 = \{x_{10}, x_{11}\}$

A set X_1 is an example of *A-definable* sets.

$X_1 = \{x_2, x_3, x_8, x_9\}$

A set X_2 is an example of *roughly A-definable* sets as obtained by its approximations:

$X_2 = \{x_2, x_3, x_7, x_8, x_9, x_{11}\}$

$$\underline{A}X_2 = E_2 \cup E_4 = \{x_2, x_3, x_8, x_9\}$$

$$\overline{A}X_2 = E_2 \cup E_3 \cup E_4 \cup E_5 = \{x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}$$

A set X_3 is an example of *externally A-non-definable* sets as obtained by its approximations:

$X_3 = \{x_1, x_3, x_6, x_8, x_9, x_{10}\}$

$\underline{A}X_3 = E_4 = \{x_8,\ x_9\}$

$\overline{A}X_3 = U = \{x_1,\ x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}$

A set X_4 is an example of *internally A-non-definable* sets as obtained by its approximations:

$X_4 = \{x_2, x_5, x_{10}\}$

$\underline{A}X_4 = \phi$

$\overline{A}X_4 = E_2 \cup E_3 \cup E_5 = \{x_2, x_3, x_5,\ x_6,\ x_7,\ x_{10}, x_{11}\}$

A set X_5 is the example of *totally A-non-definable* sets.

$X_5 = \{x_2, x_4, x_6, x_9, x_{10}\}$

3.7 Accuracy of Approximation

Rough sets provide quantitative, numerical evaluation of the quality of approximation (accuracy measure) of a set $X \subseteq U$ in the approximation space *AS* = (*U*, *IND* (*A*)), using all equivalence classes of *IND* (*A*) generated by the subset of attributes $A \subseteq Q$. Let $S =< U, Q, V, f >$ be an information system, let $A \subseteq Q$ and $X \subseteq U$ determining the approximation space *AS* = (*U*, *IND* (*A*)). The *accuracy of an approximation* of a set *X* by the set of attributes *A* (shortly *accuracy* of *X*) is defined by (3.15).

$$\alpha_A(X) = \frac{cardinality(\underline{A}X)}{cardinality(\overline{A}X)} \tag{3.15}$$

It can be easily seen that if a set *X* is *A-exactly* approximated in the approximation space defined by *A*, then $\alpha_A(X) = 1$. If a set *X* is *A-roughly* approximated in *AS*, then the range of $\alpha_A(X)$ is $0 < \alpha_A(X) < 1$. The alternative accuracy of an approximation is defined by (3.16).

$$\rho_A(X) = 1 - \alpha_A(X) \tag{3.16}$$

This is called *roughness* (*A-roughness*) of a set *X*. Roughness, as opposed to accuracy, represents a degree of inexact approximation of a set *X* in the approximation space *AS* = *(U, IND(A))* defined by $A \subseteq Q$.

The accuracy of approximation $\alpha_A(X)$ has the following properties:

1. For any $A \subseteq Q$ and $X \subseteq U$, $0 \leq \alpha_A(X) \leq 1$.
2. *A-boundary* region of *X*, $BN_A(X) = \phi$ ($\underline{A}X = \overline{A}X$ and the set *X* is *A-definable*) iff $\alpha_A(X) = 1$.
3. *A-boundary* region of *X*, $BN_A(X) \neq \phi$ (the set *X* is *A-non-definable*) iff $\alpha_A(X) < 1$.

A vague concept description can contain boundary-line objects from a universe, which cannot be with absolute certainty classified as satisfying the description of a concept. Uncertainty is related to the idea of membership of an element to a set. From rough sets perspective a set membership function can be defined, which is related to the rough sets concept. This can be considered as another numerical measure of imprecision (uncertainty). The *rough (A-rough) membership function* of an object *x* to a set *X* is defined by (3.17).

$$\mu_X^A(x) = \frac{cardinality([x]_A \cap X)}{cardinality([x]_A)} \tag{3.17}$$

where

$$0 \leq \mu^A{}_X(X) \leq 1$$

The measure characterizing a degree of uncertainty of membership of an element *x* in universe to the set *X* with respect to the possessed knowledge (in an information system) is defined by (3.18).

$$\mu_X(x) = \frac{cardinality([x]_A \cap X)}{cardinality(U)} \tag{3.18}$$

It is possible in rough sets to find a strict connection between vagueness and uncertainty. Vagueness is related to sets of objects (concepts), whereas uncertainty is related to elements of sets. Rough sets show that vagueness is defined in terms of uncertainty.

The rough set membership function can be used to define the lower and upper approximation of a set and the boundary region as in (3.19)

$$\underline{A}X = \{x \in U : \mu_X^A(x) = 1\} \tag{3.19}$$

$$\overline{A}X = \{x \in U : \mu_X^A(x) > 0\}$$

$$BN_A(X) = \{x \in U : 0 < \mu_X^A(x) < 1\}$$

[Example 3.7]

From example 3.5, for $A = \{c_1, c_2, c_3\}$ and $X_1 = \{x_2, x_3, x_5, x_6, x_7, x_9, x_{11}\}$ the accuracy of an approximation of a set X_1 by the set A is:

$$\alpha_A(X_1) = \frac{cardinality(\underline{A}X_1)}{cardinality(\overline{A}X_1)} = \frac{cardinality(\{x_2, x_3, x_5, x_6, x_7\})}{cardinality(\{x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\})} = \frac{5}{9} = 0.556$$

3.8 Approximation and Accuracy of Classification

The concept of set approximations can be extended to approximations of a classification related to the family Γ of subsets $\{X_1, X_2, \ldots, X_n\}$ from U. Let $S = < U, Q, V, f >$ be an information system, and let $A \subseteq Q$ and $\Gamma = \{X_1, X_2, \ldots, X_n\}$ for every subset $X_i \subseteq U (1 \le i \le n)$ be a classification (or a partition; a family of subsets) of U. The family of sets $\Gamma = \{X_1, X_2, \ldots, X_n\}$ is a classification in U of S, if $X_i \cap X_j = \emptyset$ for every $i,j \le n$, $i \neq j$ and $\bigcup_{i=1}^{n} X_i = U$. Sets X_is are called classes of Γ.

For $A \subseteq Q$, the *A-lower* and *A-upper approximation* of a classification of Γ on S, denoted by $\underline{A}\Gamma$ and $\overline{A}\Gamma$ respectively, are defined as follows by (3.20) and (3.21).

$$\underline{A}\Gamma = \{\underline{A}X_1, \underline{A}X_2, \ldots, \underline{A}X_n\} \tag{3.20}$$

$$\overline{A}\Gamma = \{\overline{A}X_1, \overline{A}X_2, \ldots, \overline{A}X_n\} \tag{3.21}$$

The set $\underline{A}\Gamma$ is called *A-positive* region of a classification Γ and $BN_A(\Gamma) = \overline{A}\Gamma - \underline{A}\Gamma$ is called *A-boundary* region of a classification Γ. The *A-positive* region of a classification Γ with respect to A is defined by (3.22).

$$POS_A(\Gamma) = \bigcup_{X_i \in \Gamma} \underline{A}X_i \tag{3.22}$$

A union of boundary regions of a classification *Γ* with respect to *A* is called *A-doubtful* region of a classification *Γ* in *S* as defined by (3.23).

$$BNA(\Gamma) = \bigcup_{X_i \in \Gamma} BN_A(X_i) \tag{3.23}$$

There is no *A-negative* region of a classification *Γ* in *S*, because $U = \bigcup_{X_i \in \Gamma} \overline{A}X_i$.

A classification *Γ* is called *A-definable* iff every class $X_i \in \Gamma$ is *A-definable*; otherwise the classification is called *A-non-definable*. And the classification *Γ* is called *roughly A-definable iff* $\exists X_i \in \Gamma, \underline{A}X_i \neq \phi$.

The *accuracy of approximation of a classification* by the set of attributes *A*, or *accuracy of a classification*, is defined by (3.24).

$$\alpha_A(\Gamma) = \frac{\sum_{i=1}^{n} cardinality(\underline{A}X_i)}{\sum_{i=1}^{n} cardinality(\overline{A}X_i)} \tag{3.24}$$

The *quality of approximation of a classification* by *A*, or *quality of a classification*, is defined by (3.25).

$$\rho_A(\Gamma) = \frac{\sum_{i=1}^{n} cardinality(\underline{A}X_i)}{cardinality(U)} \tag{3.25}$$

This represents a ratio of all *A-correctly* classified objects to all objects in the system *S*.

The idea of accuracy of a classification allows us to define how close one can approximate a partition (classification B^*) generated by a set of attributes $B \subseteq Q$ by another partition A^* generated by a set of attributes $A \subseteq Q$. The accuracy of approximation of classification B^* by A^* can be defined by (3.26). The following inequality holds $0 \leq \rho_A(B^*) \leq 1$ for every $A, B \subseteq Q$.

$$\rho_A(B^*) = \frac{cardinality(POS_A(B^*))}{cardinality(U)} \tag{3.26}$$

where

$$POS_A(B^*) = \bigcup_{X_i \in B^*} \overline{A}X_i \text{ : a classification } \Gamma \text{ with respect to } A \tag{3.27}$$

[Example 3.8]

From the information system MEDICAL, suppose that there is a classification $\Gamma=\{Y_1, Y_2, Y_3, Y_4\}$ where $Y_1 = \{x_2, x_4, x_5\}$, $Y_2 = \{x_1, x_3, x_6, x_7\}$, $Y_3 = \{x_8, x_9, x_{10}\}$, $Y_4 = \{x_{11}\}$.

The accuracy and the quality of a classification Γ is

$$\underline{A}\Gamma = \{\underline{A}Y_1, \underline{A}Y_2, \underline{A}Y_3, \underline{A}Y_4\} = \{\phi, \phi, E_4, \phi\} = \{\phi, \phi, \{x_8, x_9\}, \phi\}$$

$$\overline{A}\Gamma = \{\overline{A}Y_1, \overline{A}Y_2, \overline{A}Y_3, \overline{A}Y_4\} = \{\{E_1, E_2, E_3\}, \{E_1, E_2, E_3\}, \{E_4, E_5\}, \{E_5\}\} =$$
$$= \{\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \{x_8, x_9, x_{10}, x_{11}\}, \{x_{10}, x_{11}\}\}$$

$$\alpha_A(\Gamma) = \frac{0+0+2+0}{7+7+4+2} = \frac{2}{20} = 0.1$$

$$\rho_A(\Gamma) = \frac{2}{11} = 0.18$$

Therefore, it can be said that the accuracy of this classification is very poor and the classification process has to be improved towards higher accuracy.

3.9 Classification and Reduction

In many applications a classification of objects is one of the most frequently encountered tasks. Classification can be considered as a process of determining a unique class for a given object. A given set of objects, characterized by the set of condition and decision attributes, can be classified into a disjoint family of classes with respect to values of decision attributes. Each class can be determined in terms of features of corresponding condition attributes belonging to a class. If a given set of objects with a given set of attributes is classifiable, a classification may be possibly achieved by some subsets of attributes. Frequently only a few important attributes are sufficient to classify objects. This is consistent with human perception and classification ability based on intelligent attention, and selection of most important features of objects. Some attributes in an information system may be redundant and can be eliminated without losing the essential classificatory information. The process of finding a smaller set of attributes than the original one with the same or the closest classificatory power as the original set is called *attribute reduction*. As a result the original larger information system may be reduced to a smaller system containing fewer attributes.

Rough sets allow us to determine for a given information system the most important attributes from a classificatory point of view. A *reduct* is the essential part of an information system related to a subset of attributes that can discern all objects discernible by the original information system. A *core* is a common part of all reducts. Core and reduct are fundamental rough sets concepts which can be used for knowledge reduction. Some attributes may depend on each other. A change of a given attribute may cause changes of other attributes in some non-linear ways. Rough sets determine a degree of attributes' dependency and their significance. In an indiscernibility relation, a dependency of attributes is one of the important features of information systems.

Given an information system $S = < U, Q, V, f >$, with condition and decision attributes $Q = C \cup D$, for a given set of condition attributes $A \subset C$, the *A-positive* region POS_A (D) in the relation *IND (D)* can be defined by (3.28).

$$POS_A(D) = \bigcup \{\underline{A}X \mid X \in IND(D)\} \tag{3.28}$$

The positive region POS_A (D) contains all objects in U which can be classified perfectly without error into distinct classes defined by *IND* (D), based only on information in relation *IND* (A). The definition of the positive region can be formed for any two subsets of attributes, $A, B \subset Q$ in the information system S. It is known that the subset of attributes $B \subset Q$ defines the indiscernibility relation *IND* (B) and thus the classification B^* (U / *IND* (B)) with respect to the subset. The *A-positive region* of B is defined by (3.29). The A-positive region of B contains all objects that, by using attributes A, can be certainly classified to one of distinct classes of the classification B^*.

$$POS_A(B) = \bigcup_{X \in B} \underline{A}X \tag{3.29}$$

The rough sets define a degree of dependency for sets of attributes. The cardinality of the A-positive region of B is used to define a measure $\gamma_A(B)$ called a degree of dependency of the set of attributes B on A in (3.30). It can be said that the set of attributes B depends on the set of attributes A to the degree $\gamma_A(B)$.

$$\gamma_A(B) = \frac{cardinality(POS_A(B))}{cardinality(U)} \quad (3.30)$$

Suppose an information system $S = < U, Q, V, f >$ and two sets of attributes $A, B \subseteq Q$. A set of attributes B *depends on* a set A in S, denoted by $A \rightarrow B$, iff an equivalence relation satisfies $IND\ (A) \subseteq IND\ (B)$. The sets A and B are independent in S iff neither
$A \rightarrow B$ nor
$B \rightarrow A$ holds. A set B is dependent to a *degree k* on the set A in S, as denoted in (3.31).

$$A \xrightarrow{k} B,\ 0 \leq k \leq 1,\ \text{if } k = \gamma_A(B) \quad (3.31)$$

where

$\gamma_A(B)$: a degree of dependency of a set of attributes B on A

If k = 1 a set B is *totally dependent* on A (or $B \rightarrow A$), if k = 0 a set B is *totally independent* on A and otherwise a set B is *roughly dependent* on A.

A level of significance of attributes from a set A with respect to the classification B^* (U / IND (B)) generated by IND (B) may be different. The *measure of significance* (*co-efficient of significance*) of the attribute $a \in A$ from the set A with respect to the classification B^* (U / IND (B)) generated by a set B is defined by (3.32).

$$\mu_{A,B}(a) = \frac{cardinality(POS_A(B)) - cardinality(POS_{A-\{a\}}(B))}{cardinality(U)} \quad (3.32)$$

The significance of the attribute a in the set $A \subseteq Q$ computed with respect to the original classification Q^* generated by the entire set of attributes Q from the information system S is denoted as $\mu_A(a) = \mu_{A,Q}(a)$.

These are the properties of an attribute set A in an information system S as follows.

1. A set $A \subset Q$ is *dependent* in S iff $\exists B \subset A$ such that IND (B) = IND (A).
   $(\alpha_B(X) = \alpha_A(X))$

2. A set $A \subset Q$ is *independent* in *S* iff $\forall B \subset A$ such that *IND* (*B*) $\supset$ *IND* (*A*).
   $(\alpha_B(X) < \alpha_A(X))$
3. A set $A \subset Q$ is *superfluous* in *Q* iff *IND (Q-A)* = *IND* (*Q*).
   $(\alpha_{Q-A}(X) = \alpha_A(X))$
4. A set $A \subseteq Q$ is *a reduct* of *Q* in *S* iff *Q-A* is *superfluous* in *Q* and *A* is *dependent* in *S*.

A given information system may have many different reducts. If for a given information system $S = < U, Q, V, f >$, a subset $A \subset Q$ is a reduct, then the corresponding information system $S' = < U, A, V, f' >$ with the attribute set equal to a reduct *A*, is called a *reduced system* (where *f'* is the restriction of a function *f* to a set $U \times A$). In other words, a reduced system *S'* is constructed from the original system *S* by removing columns related to attributes not included in a reduct *A*.

[Example 3.9]

For an information system MEDICAL, suppose there are two subsets of attributes, $A = \{c_1, c_2, c_3\}$, $B = \{c_3\}$.

The partition A^*, which is a classification *U / IND* (*A*) related to the equivalence relation *IND* (c_1, c_2, c_3), is $A^* = \{Y_1, Y_2, Y_3, Y_4, Y_5\} = \{ \{x_1, x_4\}, \{x_2, x_3\}, \{x_5, x_6, x_7\}, \{x_8, x_9\}, \{x_{10}, x_{11}\} \}$.

In regard to the partition B^*, which is classification *U / IND*(*B*) corresponding to the equivalence relation *IND* (c_3), is $B^* = \{Z_1, Z_2\} = \{ \{x_1, x_2, x_3, x_4\}, \{x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\} \}$.

The A-positive region of *B* is

$$POS_A(B) = \bigcup_{Z \in B^*} \underline{A}Z = \underline{A}Z_1 + \underline{A}Z_2 = \{x_1, x_2, x_3, x_4\} + \{x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}.$$

A degree of dependency of the set of attributes *B* on *A* is

$$\gamma_A(B) = \frac{cardinality(POS_A(B))}{cardinality(U)} = \frac{11}{11} = 1.0$$

Therefore, it can be said that the set of attributes *B* is totally dependent on a set *A*, $A \xrightarrow{1.0} B$.

For a given original information system some attributes may be redundant with respect to a specific classification *A** generated by a set of attributes $A \subset Q$. It means that an information system may be overloaded by this redundant information. The classifiers defined for overloaded information systems may exhibit a poor generalization for new unseen objects. By virtue of the dependency properties of attributes, we can find a reduced set of the attributes, by removing superfluous attributes, without a loss of classification power of the reduced information system. It can lead to the substantial reduction of an information system by finding the optimal set of attributes sufficient for a robust classification with a higher degree of generalization.

For an information system *S* and a subset of attributes $A \subset Q$, an attribute $a \in A$ is called *dispensable* in a set *A* if *IND* (*A*) = *IND* (*A*-{*a*}), which means that indiscernibility relations generated by sets *A* and *A-{a}* are identical. Otherwise a parameter *a* is *indispensable* in *A*. It appears that the dispensable attribute does not improve the classification of the original information system *S*. It can be said that the absence of the dispensable attribute does not reduce the classificatory power of an information system and does not change the dependency relationship of the original system. On the other hand, the indispensable attributes carry the essential information about objects of an original information system, and cannot be removed without changing the classificatory power of the original system.

The set of all indispensable attributes in a set $A \subset Q$ is called a *core* of a set *A* in *S* and it is denoted by *CORE (A)*. The core contains all attributes that cannot be removed from the set *A* without losing the original classification A^*.

Consider two subsets of attributes $A, B \subset Q$ in *S*. An attribute *a* is called *B-dispensable* in the set *A* if $POS_A(B) = POS_{A-\{a\}}(B)$. Otherwise the attribute *a* is *B-indispensable*. If every attribute of *A* is *B-indispensable*, then *A* is indispensable with respect to *B*. The set of all *B-indispensable* attributes from the set *A* is called *B-relative core (or B-core)* of *A* and denoted by $CORE_B(A)$ as defined in (3.33).

$$CORE_B(A) = \{a \in A : POS_A(B) \neq POS_{A-\{a\}}(B)\} \tag{3.33}$$

A set $A \subset Q$ is called *orthogonal* if all its attributes are indispensable. A proper subset $E \subset A$ is defined as a *reduct* set of *A* in *S* if *E* is orthogonal and preserves the classification generated by *A*. Thus a reduct set of *A*, denoted by *RED (A)*, is defined by (3.34).

$$E = RED(A) \Leftrightarrow (E \subset A, IND(E) = IND(A), E - orthogonal) \quad (3.34)$$

All reducts, or a family of reducts, of a set *A* are denoted by $RED^F(A)$. The intersection of all reducts of a set *A* is a core of *A* as defined in (3.35).

$$CORE(A) = \bigcap RED(A) \quad (3.35)$$

[Example 3.10]

For a given information system MEDICAL, suppose there are two reducts B_1 and B_2 of the set of condition attributes $C = \{c_1, c_2, c_3\}$ with respect to the decision attribute $D = \{d\}$ as follows: $B_1 = \{c_1, c_2\}$, $B_2 = \{c_2, c_3\}$.

Then the core of the set of attributes B_1 and B_2 is obtained as follows.

$$CORE_D(C) = B_1 \bigcap B_2 = \{c_2\}$$

It can be said that the c_2 attribute is the most significant attribute and B_1 and B_2 are the sets of attributes that discriminate the decision attributes.

By choosing a reduct B_1, for example, the reduced decision table can be obtained by simply removing the superfluous attribute c_3 as shown in Table 3.5. The reduced decision table has the same information as the original one from the point of view of classificatory power.

Table 3.5 A reduced MEDICAL decision table

Object	*Attributes*		
	C (medical diagnoses)		*D* (disease class)
U	*c1*	*c2*	*d*
x_1	0	L	0
x_2	0	H	1
x_3	0	H	1
x_4	0	L	0
x_5	1	H	1
x_6	1	H	1
x_7	1	H	1
x_8	2	L	0
x_9	2	L	1
x_{10}	2	H	0
x_{11}	2	H	1

3.10 Decision Rules

One of the important applications of rough sets is a generation of decision rules for a given information system for a classification of known objects, or a prediction of classes for new objects unseen during design. Using an original or a reduced decision table, one can find rules classifying objects through determining the decision attribute value based on values of condition attributes.

Let $DT = < U, C \cup D, V, f >$ be a decision table with C as a set of condition attributes and D as a set of decision attributes. A decision table DT can be classified as follows:

1. *DT* is *deterministic* iff D depends on C, $C \Rightarrow D; \rho_C(D^*) = 1$.
2. *DT* is *roughly deterministic* iff D depends on C, $0 < \rho_C(D^*) < 1$.
3. *DT* is *totally non-deterministic* iff D does not dependent on C, $\rho_C(D^*) = 0$.

If *DT* is *deterministic*, a set of condition attributes *C* discriminates a set of decision attributes *D*. If *DT* is *roughly deterministic*, *D* depends on *C*, but *C* cannot discriminate *D*. If *DT* is *totally non-deterministic*, *C* is not related to *D*.

For a *deterministic* decision table, unique decisions can be determined when some conditions are satisfied (attributes taking certain values). Conversely, for a *roughly-deterministic* decision table, decisions are not uniquely determined by the conditions. For

a *non-deterministic* decision table, a subset of decisions is defined, which can be taken for specific conditions. This kind of situation is interpreted as inconsistency or uncertainty in the decision table, and thus decisions determined by the decision table are not well-defined. The properties characterizing dependency of attributes can be applied to test whether a given decision table is deterministic or non-deterministic. The notion of a reduct can be used to reduce the original decision table while preserving its classificatory power. This may lead to a design of robust classifiers with better generalization ability.

Decision rules can be derived from a decision table *DT*. Let $C^* = \{X_1, X_2, \ldots, X_r\}$ and $D^* = \{Y_1, Y_2, \ldots, Y_l\}$ be a *C*-definable and a *D*-definable classification of *U*. A class Y_i from a classification D^* can be identified with the decision i $(i=1,2,\ldots,l)$, denoted also by r_{ij}. A set of decision rules r_{ij} for all *D*-definable sets Y_j is defined by (3.36).

$$\{r_{ij}\} = \{Des_C(X_i) \Rightarrow Des_D(Y_j) : X_i \cap Y_j \neq \phi,\ for\ X_i \in C^*, Y_j \in D^*\} \tag{3.36}$$

where

$Des_C(X_i)$,$Des_D(Y_j)$: unique descriptions of classes X_i and Y_j, respectively

The decision rules r_{ij} are logically described as follows: *IF (a set of conditions) THEN (a set of decisions).*

A rule r_{ij} is said to be *deterministic* iff $X_i \subseteq Y_j$ $(X_i \cap Y_j = X_i, i = 1,2,...,r)$ in a decision table *DT*, which means $C \rightarrow D$, otherwise a rule is *non-deterministic*. In other words, if $Des_C(X_i)$ uniquely implies $Des_D(Y_j)$, then the rule r_{ij} is *deterministic*; otherwise r_{ij} is *non-deterministic*. The set of decision rules for all classes Y_j generated by a set of decision attributes *D* (*D-definable* classes in *S*) is called a *decision algorithm* resulting from the information system *S*.

[Example 3.11]

Consider the decision table in Table 3.4 with the decision attribute $D = \{d\}$, $V_d = \{0, 1\}$. The resulting partition $D^* = \{Y_1, Y_2\} = \{\{x_2, x_3, x_5, x_6, x_7, x_9, x_{11}\}, \{x_1, x_4, x_8, x_{10}\}\}$ for $Des_D(Y_1)=(d=1)$ and $Des_D(Y_2)=(d=0)$. If a reduct $A = \{c_1, c_2\}$ of the condition attribute is considered, a partition of the universe *U* corresponding to the equivalence relation *IND(A)* can be determined as below.

$$A^* = U / IND(A) = \{X_1, X_2, X_3, X_4, X_5\} = \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_5, x_6, x_7\}, \{x_8, x_9\}, \{x_{10}, x_{11}\}\}.$$

The unique descriptions of the classes X_is on the set *A* are:

$Des_A(X_1) = (c_1 = 0, c_2 = L)$

$Des_A(X_2) = (c_1 = 0, c_2 = H)$

$Des_A(X_3) = (c_1 = 1, c_2 = H)$

$Des_A(X_4) = (c_1 = 2, c_2 = L)$

$Des_A(X_5) = (c_1 = 2, c_2 = H)$.

Firstly, decision rules for the class Y_1 (sick; d=1) can be designed as follows.

Calculation	Decision rules	Decision rules logically written
$X_1 \cap Y_1 = \phi$		
$X_2 \cap Y_1 = \{x_2, x_3\}$	$r_{21} \Rightarrow Des_D(Y_1)$	$IF\ (c_1\ is\ 0)\ AND\ (c_2\ is\ H)\ THEN\ (d=1, sick)$
$X_3 \cap Y_1 = \{x_5, x_6, x_7\}$	$r_{31} \Rightarrow Des_D(Y_1)$	$IF\ (c_1\ is\ 1)\ AND\ (c_2\ is\ H)\ THEN\ (d=1, sick)$
$X_4 \cap Y_1 = \{x_9\}$	$r_{41} \Rightarrow Des_D(Y_1)$	$IF\ (c_1\ is\ 2)\ AND\ (c_2\ is\ L)\ THEN\ (d=1, sick)$
$X_5 \cap Y_1 = \{x_{11}\}$	$r_{51} \Rightarrow Des_D(Y_1)$	$IF\ (c_1\ is\ 2)\ AND\ (c_2\ is\ H)\ THEN\ (d=1, sick)$

Next, the decision rules for the class Y_2 (no disease; $d=0$) can be obtained as below.

Calculation	Decision rules	Decision rules logically written
$X_1 \cap Y_2 = \{x_1, x_4\}$	$r_{12} \Rightarrow Des_D(Y_2)$	$IF\ (c1\ is\ 0)\ AND\ (c2\ is\ L)\ THEN\ (d=0, no\ disease)$
$X_2 \cap Y_2 = \phi$		
$X_3 \cap Y_2 = \phi$		
$X_4 \cap Y_2 = \{x_8\}$	$r_{42} \Rightarrow Des_D(Y_2)$	$IF\ (c1\ is\ 2)\ AND\ (c2\ is\ L)\ THEN\ (d=0, no\ disease)$
$X_5 \cap Y_2 = \{x_{10}\}$	$r_{52} \Rightarrow Des_D(Y_2)$	$IF\ (c1\ is\ 2)\ AND\ (c2\ is\ H)\ THEN\ (d=0, no\ disease)$

Therefore, the entire decision algorithm is a set of decision rules {r_{21}, r_{31}, r_{41}, r_{51}, r_{12}, r_{42}, r_{52}} for both classes.

Chapter 4

4. Rough-Fuzzy Hybridization

Fuzzy and rough sets theories are generalizations of set theory in mathematics to describe vagueness, uncertainty, and imprecision. A characteristic function of a fuzzy set uses a degree of membership in [0, 1], whereas a characteristic function of a rough set employs three membership functions; a reference set, its lower and upper approximations in an approximation space. There have been extensive theoretical contributions on the relationships between rough sets and fuzzy sets [54], [55], [56] and many approaches have been proposed on the combination of rough and fuzzy sets [57], [58], [59]: rough-fuzzy sets, fuzzy-rough sets.

The main objective of this chapter is to review theoretical approaches on combination of rough and fuzzy sets by utilizing α-level set method, which is based on relationships between rough sets and fuzzy sets. Most of the mathematical symbols for representing fuzzy sets and rough sets are identical as defined in Chapter 2 and 3, respectively.

4.1 Introduction

In particular, a rough-fuzzy set is defined as an approximation of a fuzzy set in a crisp approximation space, while a fuzzy-rough set is defined as an approximation of a crisp set in a fuzzy approximation space. In generalization, the category of an approximation can be interpreted in these three different areas; a family of rough sets, a family of rough- fuzzy sets, and a family of fuzzy-rough sets. The approximation of a fuzzy set in a fuzzy approximation space leads to a more general framework.

By definition, analysis, and operation of a set with fuzzy concepts, it is simpler to utilize a set-method, for instance, the use of α-level sets of a fuzzy set. One example of using a set-method on combination of rough and fuzzy sets is a more general framework suggested by Klir and Yuan [60].

4.2 Fuzzy Sets

Fuzzy sets are a generalization of sets in which their membership functions are defined in [0, 1] of real number domain.
The α-level set, or α-cut, of a fuzzy set F is defined by (4.1).

$$F_\alpha = \{x \in U \mid \mu_F(x) \geq \alpha\} \tag{4.1}$$

where

U: a universe

x: an element in U

F: a fuzzy set on U

μ_F: a membership function of a fuzzy set F on U

α: a real number in [0, 1]

A fuzzy set determines a family of nested subsets of U through α-cuts. On the other hand, a fuzzy set F can be re-constructed from its α-level sets as defined by (4.2).

$$\mu_F(x) = \sup\{\alpha \mid x \in F_\alpha\} \tag{4.2}$$

The equality and inclusion of two fuzzy sets, F_1 and F_2 can be represented by (4.3).

$$\begin{aligned} F_1 = F_2 \quad &\Leftrightarrow \quad \mu_{F_1}(x) = \mu_{F_2}(x) \quad \forall x \in U \\ F_1 \subseteq F_2 \quad &\Leftrightarrow \quad \mu_{F_1}(x) \leq \mu_{F_2}(x) \quad \forall x \in U \end{aligned} \tag{4.3}$$

Employing of α-level sets, equations in (4.3) can be equivalently defined by (4.4).

$$\begin{aligned} F_1 = F_2 \quad &\Leftrightarrow \quad F_{1\alpha} = F_{2\alpha} \quad \forall \alpha \in [0,1] \\ F_1 \subseteq F_2 \quad &\Leftrightarrow \quad F_{1\alpha} \subseteq F_{2\alpha} \quad \forall \alpha \in [0,1] \end{aligned} \tag{4.4}$$

Thus, either definition of fuzzy sets can be used. One of main advantages of these set-based representations is that it establishes a linkage between fuzzy sets and sets, which shows the inherent structure of a fuzzy set.

Utilizing the standard max-min system proposed by Zadeh [1], the fuzzy-set complement, intersection, and union are defined by (4.5).

$$\mu_{\neg F}(x) = 1 - \mu_F(x) \tag{4.5}$$

$$\mu_{F_1 \cap F_2}(x) = \min[\ \mu_{F_1}(x),\ \mu_{F_2}(x)\ ]$$

$$\mu_{F_1 \cup F_2}(x) = \max[\ \mu_{F_1}(x),\ \mu_{F_2}(x)\ ]$$

where

F_1, F_2: two fuzzy sets defined in a universe U

$\mu_{\neg F}(x)$: a membership function of the complement of a fuzzy set F

$\mu_{F_1 \cap F_2}(x)$, $\mu_{F_1 \cup F_2}(x)$: membership functions of sets of the intersection, union of F_1 and F_2, respectively

Using α-level sets, equations in (4.5) are represented by (4.6).

$$(\neg F_1)_\alpha = \neg F_{(1-\alpha)^+}$$

$$(F_1 \cap F_2)_\alpha = F_{1\alpha} \cap F_{2\alpha} \tag{4.6}$$

$$(F_1 \cup F_2)_\alpha = F_{1\alpha} \cup F_{2\alpha}$$

where

$F_{\alpha+}$: a strong α-cut of a fuzzy set F, which is defined by (4.7)

$$F_{\alpha+} = \{x \in U \mid \mu_F(x) > \alpha\} \tag{4.7}$$

An important feature of fuzzy set operations is that they are truth-functional [21]. The membership functions of the complement, intersection, and union of fuzzy sets can be obtained, which is based only on the membership functions of the fuzzy sets involved.

4.3 Rough Sets

Given an information system $S = < U, Q, V, f >$ and $A \subseteq Q$ as a subset of attributes. For any given subset of attributes $A \subseteq Q$, the *indiscernibility relation* on *A, IND (A)* denoted by $\tilde{A}$, is an *equivalence relation* on universe *U*. For a given arbitrary set $X \subseteq U$, it may or may not be possible to describe the arbitrary set *X* exactly in its approximation space $(U, \tilde{A})$. In rough sets theory, an arbitrary set in a universe can be represented or characterized by its lower and upper approximations as defined by (4.8). The pair ($\underline{A}X$, $\overline{A}X$) is called a rough set on U with a reference set X.

$$
\begin{aligned}
&\textit{reference set}: \quad X \subseteq U \\
&\textit{lower approximation}: \quad \underline{A}X = \{x \in U \mid [x]_A \subseteq X\} \\
&\textit{upper approximation}: \quad \overline{A}X = \{x \in U \mid [x]_A \cap X \neq \phi\}
\end{aligned}
\tag{4.8}
$$

The characteristic functions of $\underline{A}X$ and $\overline{A}X$ are called strong and weak membership functions [55]. The physical meaning of lower and upper approximations may be understood better by the following two representations in (4.9) and (4.10).

$$
\begin{aligned}
\mu_{\underline{A}X}(x) &= \inf\{\mu_X(y) \mid y \in U, (x,y) \in \tilde{A}\} \\
\mu_{\overline{A}X}(x) &= \sup\{\mu_X(y) \mid y \in U, (x,y) \in \tilde{A}\}
\end{aligned}
\tag{4.9}
$$

$$
\begin{aligned}
\mu_{\underline{A}X}(x) &= \inf\{1 - \mu_{\tilde{A}}(x,y) \mid y \notin X\} \\
\mu_{\overline{A}X}(x) &= \sup\{\mu_{\tilde{A}}(x,y) \mid y \in X\}
\end{aligned}
\tag{4.10}
$$

The weak and strong membership functions of a rough set can be computed from the membership function of the reference set if the equivalence relation is used to select elements to be considered. Alternatively, they can also be computed from the membership functions of the equivalent relation if the reference set is used to select elements to be considered. These two views are important on the combination of rough and fuzzy sets. For convenience, the strong and weak membership functions of a rough set can be represented by (4.11).

$$
\begin{aligned}
\mu_{\underline{A}X}(x) &= \inf\{\max[\mu_X(y), 1 - \mu_{\tilde{A}}(x,y)] \mid y \in U\} \\
\mu_{\overline{A}X}(x) &= \sup\{\min[\mu_X(y), \mu_{\tilde{A}}(x,y)] \mid y \in U\}
\end{aligned}
\tag{4.11}
$$

Rough sets are monotonic with respect to set inclusion as shown by (4.12).

$$
\begin{aligned}
X_1 \subseteq X_2 \quad &\Rightarrow \quad \underline{A}X_1 \subseteq \underline{A}X_2 \\
X_1 \subseteq X_2 \quad &\Rightarrow \quad \overline{A}X_1 \subseteq \overline{A}X_2
\end{aligned}
\tag{4.12}
$$

where

X_1 and X_2: arbitrary sets in *U*

Let $\tilde{A}_1$ and $\tilde{A}_2$ be two equivalence relations on *U*. The first equivalence relation $\tilde{A}_1$ is a refinement of $\tilde{A}_2$, or $\tilde{A}_2$ is a coarsening of $\tilde{A}_1$, if $\tilde{A}_1 \subseteq \tilde{A}_2$. A refinement relation further divides the equivalence classes of a coarsening one. That is, $\tilde{A}_1$ is a refinement of $\tilde{A}_2$ *iff* $[x]_{\tilde{A}_1} \subseteq [x]_{\tilde{A}_2} \quad \forall x \in U$. The finest equivalence relation is the identity relation, whereas the coarsest relation is the Cartesian product *U* × *U*.

Rough sets are monotonic with respect to refinement of equivalence relations. If an equivalence relation $\tilde{A}_1$ is a refinement of another equivalence relation $\tilde{A}_2$, for any $X \subseteq U$, the property of equivalence relation with respect to set inclusion is shown by (4.13).

$$\begin{aligned} \tilde{A}_1 \subseteq \tilde{A}_2 \quad &\Rightarrow \quad \underline{A_1} X \supseteq \underline{A_2} X \\ \tilde{A}_1 \subseteq \tilde{A}_2 \quad &\Rightarrow \quad \overline{A_1} X \subseteq \overline{A_2} X \end{aligned} \tag{4.13}$$

Approximation of a set in an approximation space refined is more accurate in the sense that both lower and upper approximations are closer to the given set. The two monotonic properties of rough sets are useful to the combination of rough and fuzzy sets.

4.4 Combination of Rough and Fuzzy Sets

There have been different proposals of rough-fuzzy sets and fuzzy-rough sets for defining such terms mathematically. The main results are briefly reviewed before presenting some of their analysis.

Rough-Fuzzy Sets defined by Dubois and Prade deal with the approximation of fuzzy sets in an approximation space [58]. Given a fuzzy set *F*, the result of approximation is a pair of fuzzy sets on the equivalence class, *U / IND (A)* as defined by (4.14).

$$\begin{aligned} \mu_{\underline{A}F}([x]_A) &= \inf\{\mu_F(y) \mid y \in [x]_A\} \\ \mu_{\overline{A}F}([x]_A) &= \sup\{\mu_F(y) \mid y \in [x]_A\} \end{aligned} \tag{4.14}$$

where

U: a universe

A: a subset of attributes $A \subseteq Q$ in a given information system S

IND (A), or $\tilde{A}$: an indiscernibility relation, or an equivalence relation on U

$\underline{A}F$: a lower approximation of a given fuzzy set F

$\overline{A}F$: a upper approximation of a given fuzzy set F

$[x]_A$: an equivalence class which contains x on an equivalence relation *IND (A)*

y: an element belongs to $[x]_A$ in U

Using the extension principle, the pair can be extended to a pair of rough sets on the universe U as defined by (4.15).

$$\begin{aligned} \mu_{\underline{A}F}(x) &= \inf\{\mu_F(y) \mid y \in [x]_A\} \\ \mu_{\overline{A}F}(x) &= \sup\{\mu_F(y) \mid y \in [x]_A\} \end{aligned} \quad (4.15)$$

This pair can be represented in another way by expressing rough sets using the characteristic functions of lower and upper approximation as defined by (4.16).

$$\begin{aligned} \mu_{\underline{A}F}(x) &= \inf\{\max[\mu_F(y), \mu_{IND(A)}(x,y)] \mid y \in U\} \\ \mu_{\overline{A}F}(x) &= \sup\{\min[\mu_F(y), 1-\mu_{IND(A)}(x,y)] \mid y \in U\} \end{aligned} \quad (4.\ 16)$$

The pair ($\underline{A}F$, $\overline{A}F$) is called a rough-fuzzy set on U with reference fuzzy set F.

Fuzzy-Rough Sets defined by Dubois and Prade [59] are originated from Waillaeys and Malvache [61] for defining a fuzzy set with respect to a family of fuzzy sets. It deals with the approximation of fuzzy sets in a fuzzy approximation space defined by a fuzzy similarity relation or defined by a fuzzy partition. The results for fuzzy-rough sets reviewed here are based on a fuzzy similarity relation. A fuzzy similarity relation $\tilde{R}$ is a fuzzy subset of $U \times U$ and has three properties defined by (4.17).

$$\begin{aligned} &\textit{reflexivity}: && \forall x \in U,\ \mu_{\tilde{R}}(x,x) = 1 \\ &\textit{symmetry}: && \forall x,y \in U,\ \mu_{\tilde{R}}(x,y) = \mu_{\tilde{R}}(y,x) \\ &\textit{transitivity}: && \forall x,y,z \in U,\ \mu_{\tilde{R}}(x,z) \geq \min[\mu_{\tilde{R}}(x,y), \mu_{\tilde{R}}(y,z)] \end{aligned} \quad (4.17)$$

Given a fuzzy similarity relation $\tilde{R}$, the pair $(U, \tilde{R})$ is called a fuzzy approximation space. A fuzzy similarity relation can be used to define a fuzzy partition of the universe. A fuzzy equivalence class [*x*] of elements close to *x* is defined by (4.18).

$$\mu_{[x]_{\tilde{R}}}(y) = \mu_{\tilde{R}}(x, y) \tag{4.18}$$

A family of all fuzzy equivalence classes is denoted by $U / \tilde{R}$. For a fuzzy set *F*, its approximation in the fuzzy approximation space is called a fuzzy-rough set, which is a pair of fuzzy sets on $U / \tilde{R}$ as defined by (4.19).

$$\begin{aligned} \mu_{\underline{\tilde{R}}F}([x]_{\tilde{R}}) &= \inf\{\ \max[\ \mu_F(y), 1 - \mu_{[x]_{\tilde{R}}}(y)\ ] \mid y \in U\} \\ \mu_{\overline{\tilde{R}}F}([x]_{\tilde{R}}) &= \sup\{\ \min[\ \mu_F(y), \mu_{[x]_{\tilde{R}}}(y)\ ] \mid y \in U\} \end{aligned} \tag{4.19}$$

The pair can be extended to a pair of fuzzy sets on the universe *U* as defined by (4.20).

$$\begin{aligned} \mu_{\underline{\tilde{R}}F}(x) &= \inf\{\ \max[\ \mu_F(y), 1 - \mu_{\tilde{R}}(x, y)\ ] \mid y \in U\} \\ \mu_{\overline{\tilde{R}}F}(x) &= \sup\{\ \min[\ \mu_F(y), \mu_{\tilde{R}}(x, y)\ ] \mid y \in U\} \end{aligned} \tag{4.20}$$

The approximation of a crisp set in a fuzzy approximation space may be considered as a special case. By comparing Equations (4.16) and (4.20), it can be concluded that rough-fuzzy sets are special cases of fuzzy-rough sets as defined by Dubois and Prade. Although the names of rough-fuzzy sets and fuzzy-rough sets are symmetric, the roles played by them are not symmetric.

Nakamura [62] defined a fuzzy rough set by using a family of equivalence relations induced by different level sets of a fuzzy similarity relation. Nanda and Maumdar [63] suggested a different proposal for the definition of fuzzy rough sets by extending the work of Iwinski [64]. Their definition is based on a fuzzification of the lower and upper bounds of Iwinski rough sets. It may be related to the concept of interval-valued fuzzy sets. The same definition was also used by Biswias [57]. Kuncheva [65] defined the notion of fuzzy rough sets which models the approximation of a fuzzy set based on a weak fuzzy partition. It uses measures of fuzzy set inclusion. A number of different definitions may indeed be obtained with various measures of fuzzy set inclusion.

The review shows that the same notions of rough fuzzy set and fuzzy rough sets are used with different meanings by different authors. The functional approaches clearly defined various notions mathematically. However, the physical meanings of these notions are not clearly interpreted. In the rest of this chapter, this issue will be addressed.

The approximation of a fuzzy set in a crisp approximation space is called a rough fuzzy set, to be consistent with the naming of rough set as the approximation of a crisp set in a crisp approximation space. The approximation of a crisp set in a fuzzy approximation space is called a fuzzy rough set. Such a naming scheme has been used by Klir and Yuan [60], and Yao [66]. Under this scheme, these two models are complementary to each other, in a similar way that rough sets and fuzzy sets complementary to each other. In contrast to the proposal of Dubois and Prade [59], rough fuzzy sets are not considered as special cases of fuzzy rough sets. As a result, the framework of the approximation of a fuzzy set in a fuzzy approximation space is considered to be a more general model which unifies rough fuzzy sets and fuzzy rough sets. All these notions are interpreted based on the concept of alpha-level sets, which may be useful for their successful applications.

Because most of the studies mentioned on the combination of rough and fuzzy sets are based on the functional approach (i.e., the α-level sets of a fuzzy set), the α-level set-based functional approach is used for the combination of fuzzy and rough sets.

4.4.1 Rough-Fuzzy Sets

Suppose the approximation of a fuzzy set $F = (F_\alpha)_\alpha$, $\alpha \in [0,1]$ in an approximation space (U, $IND(A)$). For each α-level set F_α, a rough set is defined by (4.21).

$$\begin{aligned} &\textit{reference set}: \quad F_\alpha \\ &\textit{lower approximation}: \quad \underline{A}F_\alpha = \{x \in U \mid [x]_A \subseteq F_\alpha\} \\ &\textit{upper approximation}: \quad \overline{A}F_\alpha = \{x \in U \mid [x]_A \cap F_\alpha \neq \phi\} \end{aligned} \tag{4.21}$$

The pair $(\underline{A}F_\alpha, \overline{A}F_\alpha)$ is a rough set with a reference set, F_α.

The use of α-level sets provides a clear interpretation of rough-fuzzy sets. A fuzzy set *F* is described by a pair of fuzzy sets in an approximation space. It lies between the lower and upper approximation. In other words, a rough-fuzzy set is characterized by three fuzzy sets defined by (4.22).

$$\begin{aligned} &\textit{reference fuzzy set}: \quad \mu_F \\ &\textit{lower approximation}: \quad \mu_{\underline{A}F}(x) = \inf\{\mu_F(y) \mid y \in U, (x,y) \in IND(A)\} \\ &\textit{upper approximation}: \quad \mu_{\overline{A}F}(x) = \sup\{\mu_F(y) \mid y \in U, (x,y) \in IND(A)\} \end{aligned} \tag{4.22}$$

An α-level set of a rough-fuzzy set is defined by (4.23) in terms of the α-level sets of a fuzzy set *F*.

$$(\underline{A}F, \overline{A}F)_\alpha = (\underline{A}F_\alpha, \overline{A}F_\alpha) = ((\underline{A}F)_\alpha, (\overline{A}F)_\alpha) \tag{4.23}$$

Rough-fuzzy sets have the following properties in (4.24) for two fuzzy sets F_1 and F_2.

i) $\underline{A}(\neg F_1) = \neg\overline{A}(F_1), \quad \overline{A}(\neg F_1) = \neg\underline{A}(F_1)$ (4.24)

ii) $\underline{A}(U) = U, \quad \overline{A}(\phi) = \phi$

iii) $\underline{A}(F_1 \cap F_2) = \underline{A}F_1 \cap \underline{A}F_2, \quad \overline{A}(F_1 \cup F_2) = \overline{A}F_1 \cup \overline{A}F_2$
$\underline{A}(F_1 \cup F_2) \supseteq \underline{A}F_1 \cup \underline{A}F_2, \quad \overline{A}(F_1 \cap F_2) \subseteq \overline{A}F_1 \cap \overline{A}F_2$

iv) $\underline{A}F_1 \subseteq F_1, \quad F_1 \subseteq \overline{A}F_1$

v) $F_1 \subseteq \underline{A}(\overline{A}(F_1)), \quad \overline{A}(\underline{A}(F_1)) \subseteq F_1$

vi) $\underline{A}F_1 \subseteq \underline{A}(\underline{A}(F_1)), \quad \overline{A}(\overline{A}(F_1)) \subseteq \overline{A}F_1$

Rough-fuzzy sets are monotonic with respect to fuzzy set inclusion as shown by (4.25).

$$\begin{aligned} F_1 \subseteq F_2 &\Rightarrow \underline{A}F_1 \subseteq \underline{A}F_2 \\ F_1 \subseteq F_2 &\Rightarrow \overline{A}F_1 \subseteq \overline{A}F_2 \end{aligned} \tag{4.25}$$

They are also monotonic with respect to refinement of equivalence relations for two equivalence relations $\tilde{A}_1$, $\tilde{A}_2$ and a fuzzy set *F*, as shown by (4.26).

$$\begin{aligned} \tilde{A}_1 \subseteq \tilde{A}_2 &\Rightarrow \underline{A}_1F \supseteq \underline{A}_2F \\ \tilde{A}_1 \subseteq \tilde{A}_2 &\Rightarrow \overline{A}_1F \subseteq \overline{A}_2F \end{aligned} \tag{4.26}$$

4.4.2 Fuzzy-Rough Sets

The concept of approximation spaces can be generalized by using fuzzy relations [67], [58]. Suppose a fuzzy approximation space (U, $\tilde{R}$), where $\tilde{R}$ is a fuzzy similarity relation. Each of $\tilde{R}$'s β-level sets is an equivalence relation [68]. The relation $\tilde{R}$ can be represented by a family of equivalence relations as defined by (4.27).

$$\tilde{R} = (\tilde{R}_\beta)_\beta, \quad \beta \in [0,1] \tag{4.27}$$

This family defines a family of approximation spaces, $(U, \tilde{R}_\beta)_\beta$.

Given a subset X of U, suppose its approximation in each of the approximation spaces. For $\beta \in [0,1]$, a rough set is defined by (4.28).

$$\begin{aligned}
&\textit{reference set}: \quad X \subseteq U \\
&\textit{lower approximation}: \quad \underline{\tilde{R}}_\beta X = \{x \in U \mid [x]_{\tilde{R}_\beta} \subseteq X\} \\
&\textit{upper approximation}: \quad \overline{\tilde{R}}_\beta X = \{x \in U \mid [x]_{\tilde{R}_\beta} \cap X \neq \phi\}
\end{aligned} \tag{4.28}$$

With respect to a fuzzy approximation space, we obtain a family of rough sets as defined by (4.29).

$$(\underline{\tilde{R}}_\beta X, \overline{\tilde{R}}_\beta X)_\beta, \quad \beta \in [0,1] \tag{4.29}$$

The pair of fuzzy sets $(\underline{\tilde{R}}X, \overline{\tilde{R}}X)$ is called a fuzzy-rough set with reference set X. A fuzzy-rough set is characterized by a crisp set and two fuzzy sets as defined by (4.30).

$$\begin{aligned}
&\textit{reference set}: \quad X \subseteq U \\
&\textit{lower approximation}: \quad \mu_{\underline{\tilde{R}}X}(x) = \inf\{1 - \mu_{\tilde{R}}(x, y) \mid y \notin X\} \\
&\textit{upper approximation}: \quad \mu_{\overline{\tilde{R}}X}(x) = \sup\{\mu_{\tilde{R}}(x, y) \mid y \in X\}
\end{aligned} \tag{4.30}$$

A β-level set of a fuzzy-rough set is defined by (4.31) in terms of the β-level sets of the fuzzy similarity relation, which is a rough set with a reference set X in the approximation space (U, $\tilde{R}_\beta$).

$$(\underline{\tilde{R}}X, \overline{\tilde{R}}X)_\beta \quad = \quad (\underline{\tilde{R}}_\beta X, \overline{\tilde{R}}_\beta X) \quad = \quad ((\underline{\tilde{R}}X)_{(1-\beta)}, (\overline{\tilde{R}}X)_\beta) \qquad (4.31)$$

Based on the properties of rough sets, properties of fuzzy-rough sets are shown by (4.32) for X_1, X_2 in U.

i) $\underline{\tilde{R}}(\neg X_1) = \neg\overline{\tilde{R}}(X_1) \quad \overline{\tilde{R}}(\neg X_1) = \neg\underline{\tilde{R}}(X_1)$ (4.32)

ii) $\underline{\tilde{R}}(U) = U \quad \overline{\tilde{R}}(\phi) = \phi$

iii) $\underline{\tilde{R}}(X_1 \cap X_2) = \underline{\tilde{R}}X_1 \cap \underline{\tilde{R}}X_2, \quad \overline{\tilde{R}}(X_1 \cup X_2) = \overline{\tilde{R}}X_1 \cup \overline{\tilde{R}}X_2$

$\underline{\tilde{R}}(X_1 \cup X_2) \supseteq \underline{\tilde{R}}X_1 \cup \underline{\tilde{R}}X_2, \quad \overline{\tilde{R}}(X_1 \cap X_2) \subseteq \overline{\tilde{R}}X_1 \cap \overline{\tilde{R}}X_2$

iv) $\underline{\tilde{R}}X_1 \subseteq X_1, \quad X_1 \subseteq \overline{\tilde{R}}X_1$

Fuzzy-rough sets are monotonic with respect to set inclusion as shown by (4.33).

$$\begin{aligned} X_1 \subseteq X_2 \quad &\Rightarrow \quad \underline{\tilde{R}}X_1 \subseteq \underline{\tilde{R}}X_2 \\ X_1 \subseteq X_2 \quad &\Rightarrow \quad \overline{\tilde{R}}X_1 \subseteq \overline{\tilde{R}}X_2 \end{aligned} \qquad (4.33)$$

They are monotonic with respect to the refinement of fuzzy similarity relations. A fuzzy similarity relation $\tilde{R}_1$ is a refinement of another fuzzy similarity relation $\tilde{R}_2$ if $\tilde{R}_1$ belongs or equal to $\tilde{R}_2$, which is a straightforward generalization of the refinement of crisp relations. The monotonicity of fuzzy-rough sets with respect to the refinement of the fuzzy similarity relation can be expressed by (4.34).

$$\begin{aligned} \tilde{R}_1 \subseteq \tilde{R}_2 \quad &\Rightarrow \quad \underline{\tilde{R}}_1 X \supseteq \underline{\tilde{R}}_2 X \\ \tilde{R}_1 \subseteq \tilde{R}_2 \quad &\Rightarrow \quad \overline{\tilde{R}}_1 X \subseteq \overline{\tilde{R}}_2 X \end{aligned} \qquad (4.34)$$

4.4.3 Approximation of Fuzzy Sets in Fuzzy Approximation Spaces

This section examines the approximation of a fuzzy set in a fuzzy approximation space. In this framework, there is a family of α-level sets, $(F_\alpha)_\alpha, \alpha \in [0,1]$, representing a fuzzy set F; whereas there is another family of β-level sets, $(\tilde{R}_\beta)_\beta, \beta \in [0,1]$, representing a fuzzy similarity relation $\tilde{R}$. Each α-level set F_α is a crisp set, and each β-

level relation $\tilde{R}_\beta$ is an equivalence relation. Therefore, rough sets, rough-fuzzy sets, and fuzzy-rough sets can be viewed as special cases of a generalized model.

For a fixed pair of number $(\alpha, \beta) \in [0,1] \times [0,1]$, a sub-model is obtained, in which a crisp set F_α is approximated in a crisp approximation space (U, $\tilde{R}_\beta$). The result is a rough set $(\underline{\tilde{R}}_\beta F_\alpha, \overline{\tilde{R}}_\beta F_\alpha)$ with the reference set F_α. For a fixed β, a sub-model is found in which a fuzzy set $(F_\alpha)_\alpha, \alpha \in [0,1]$ is approximated in a crisp approximation space (U, $\tilde{R}_\beta$). The result is a rough-fuzzy set $(\underline{\tilde{R}}_\beta F, \overline{\tilde{R}}_\beta F)$ with the reference fuzzy set F. On the other hand, for a fixed α, a sub-model is also identified in which a crisp set F_α is approximated in a fuzzy approximation space $((U, \tilde{R}_\beta))_\beta, \beta \in [0,1]$. The result is s fuzzy-rough set $(\underline{\tilde{R}} F_\alpha, \overline{\tilde{R}} F_\alpha)$ with the reference set F_α. In a generalization model, both α and β are not fixed. The result may be interpreted in three different views.

A family of rough sets:

$$(\underline{\tilde{R}}_\beta F_\alpha, \overline{\tilde{R}}_\beta F_\alpha) \qquad \alpha \in [0,1], \beta \in [0,1] \tag{4.35}$$

This represents the rough set approximation of each α -level set of a fuzzy set F in an approximation space induced by a β -level relation of a fuzzy similarity relation $\tilde{R}$. Under this interpretation, the relationships between different α -level sets of F, and the relationships between different β -level relations of $\tilde{R}$, are not taken into consideration.

A family of rough-fuzzy sets:

$$(\underline{\tilde{R}}_\beta F, \overline{\tilde{R}}_\beta F)_\beta \qquad \beta \in [0,1] \tag{4.36}$$

The second category takes into consideration the relationships between different α-level sets of a fuzzy set F. The relationships between β -level sets of a fuzzy similarity relation $\tilde{R}$ are not considered.

A family of fuzzy-rough sets:

$$(\underline{\tilde{R}} F_\alpha, \overline{\tilde{R}} F_\alpha) \qquad \alpha \in [0,1] \tag{4.37}$$

By employing the relationship between different β-level relations of a fuzzy relation $\tilde{R}$, a family of fuzzy rough sets is obtained. It does not take account the relationships between different α-level sets of a fuzzy set F.

The above interpretations depend on the ways in which the family of rough sets, $(\underline{\tilde{R}}_{\beta}F_{\alpha}, \overline{\tilde{R}}_{\beta}F_{\alpha})$, $\alpha \in [0,1]$, $\beta \in [0,1]$, are grouped. An interesting problem is how to take into consideration both relationships between different α-level sets of fuzzy sets, and the relationships between different β-level relations of fuzzy similarity relations. It can be concluded that membership functions of rough sets, rough-fuzzy sets, and fuzzy-rough sets can be computed uniformly using the same scheme defined by (4.38).

$$
\begin{aligned}
\mu_{\underline{\Gamma}(\Delta)}(x) &= \inf\{ \max[\mu_{\Delta}(y), 1-\mu_{\Gamma}(x,y)] \mid y \in U\} \\
\mu_{\overline{\Gamma}(\Delta)}(x) &= \sup\{ \min[\mu_{\Delta}(y), \mu_{\Gamma}(x,y)] \mid y \in U\}
\end{aligned}
\quad (4.38)
$$

where

Γ: a variable which takes either an equivalence relation or a fuzzy similarity relation as its value

Δ: a variable that takes either a crisp set or a fuzzy set as its value

The same scheme is used by Dubois and Prade [58] to define a pair of fuzzy sets as the result of approximating a fuzzy set in a fuzzy approximation space. This involves the combination of degrees of memberships of a fuzzy set and a fuzzy similarity relation. The physical meaning is not entirely clear. It is questionable that an element with α degree membership belonging to a fuzzy set would have the same physical interpretation as a pair with α degree membership belonging to a fuzzy relation, as the universes of the former and latter are quite different. For this reason, in this study here it is not mixed between the membership functions of a fuzzy set and a fuzzy similarity relation. As seen from equations, the *inf* and *sup* operations are performed on one membership function.

Chapter 5

5. A framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS)

In past decades, fuzzy systems have been combined popularly with neural networks for performing pattern classification tasks. Many approaches [69], [70], [71] have been proposed to address the issue of automatic generation of membership functions and fuzzy rules from input-output data sets and also subsequent adjustment of them towards more satisfactory performance. Most of these schemes that incorporate the learning property of neural networks within a fuzzy system framework have provided encouraging results. However, most of these techniques have drawbacks associated with the maximum number of resulting fuzzy rules, which increase exponentially when higher numbers of input variables are employed. As a result, the computational load required to search for a corresponding fuzzy rule becomes very heavy.

Rough set theory has recently been deployed with fuzzy inference systems to obtain more compact information from the given data and to effectively reduce the given knowledge [72]. This is an attributes-reduced information system resulting in the absolute minimal set of decision rules. Rough set theory provides a methodology to do this in data analysis based on empirical data and it has been applied to a variety of practical applications. The effective knowledge-reduction using feature reduction approaches can be applied to the existing fuzzy systems to resolve the difficulties mentioned above. The Takagi-Sugeno (T-S) type fuzzy model [7] has an ability to exactly approximate non-linear systems using a combination of linear systems. It is a very powerful tool as a universal approximator [29], [30] in non-linear system modeling.

Consequently, if a minimal set of rules generated by the rough set approach is suitable to carry out the T-S type fuzzy inference, not only the number of fuzzy antecedent variables involved but also the number of fuzzy inference rules can be effectively reduced. The advantages of both the rough set approach and the T-S fuzzy model are combined in order to introduce a new framework of Adaptive T-S type Rough-Fuzzy

Inference Systems (ARFIS). Without a loss of generality, a Multi-Input-Single-Output (MISO) type fuzzy inference system is assumed since it is known that Multi-Input-Multi-Output (MIMO) fuzzy systems can be decomposed into a number of MISO fuzzy systems [73].

A functional block diagram of the proposed framework of ARFIS is shown in Figure 5.1. As a brief overview, a pre-processing is applied on the given input and output data to generate two major components of the proposed system; 1) adaptive fuzzy clusters for the input using the Fuzzy C-Means (FCM) clustering and 2) decision rules of the given information using the decision rule generation algorithm in the rough set approach. The obtained fuzzy clusters and decision rules are used to model membership functions and T-S type fuzzy rules in the knowledge-base. Once the T-S type rough-fuzzy inference system is constructed with training data, a system evaluation process is carried out as a post-processing. If the system performance is not satisfactory, an advanced adjustment is applied to the knowledge-base towards better accuracy.

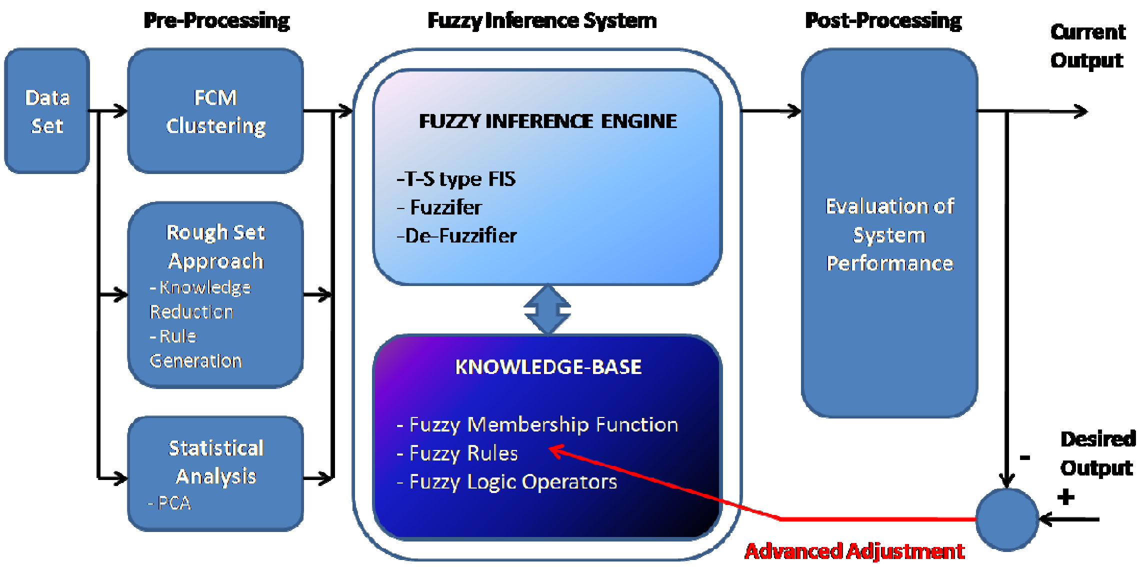


Figure 5.1 A functional block diagram of the proposed framework of ARFIS

5.1 Automatic Generation of Membership Functions

In order to build a T-S type rough-fuzzy inference system, firstly an automatic generation of membership functions is required. The Fuzzy C-Means (FCM) clustering algorithm [22] is employed to find each cluster by minimizing the FCM objective function which measures distances between data points and cluster prototypes. The FCM clustering algorithm is an unsupervised clustering method whose aim is to establish a fuzzy partition of a set of pattern vectors in *C* number of clusters and the corresponding set of cluster prototypes towards the local minimum of their objective function. An objective function *J* defined by (5.1) measures the fitting between the clusters and their cluster prototypes.

$$J_m(M, v) = \sum_{k=1}^{n} \sum_{i=1}^{C} (u_{ik})^m (d_{ik})^2 \tag{5.1}$$

where

$u_{ik} \in [0,1]$: a membership degree of the *k*-th pattern vector to the *i*-th cluster represented by its cluster prototype v_i

v : a cluster prototype of u_i

d_{ik} : *Euclidean* distance, $d_{ik} = \|x_k - v_i\|$ on R^p if a pattern vector is in a *p*-dimensional space

m : a weighting exponent so-called fuzzifier, $m \in [1, \infty)$, which makes the resulting partitions more or less fuzzy

After the FCM clustering, each membership function of the *j*-th feature, x_j, is obtained by plotting the elements of each row of the membership matrix *M* versus x_j values. Two procedures are applied for each membership function to form their shapes and to fit their membership values.

1) Finding outer shapes: Amongst all data points of each membership function after plotting the entries as above, select only the maximum membership degree for each value of the *j*-th feature, x_j. These maximum membership degrees will be used in the fitting process to generate prototypes of their corresponding membership functions as follows.

2) Removing false representation:

$$\sum_{i=1}^{C} u_{ik} = 1.0 \tag{5.2}$$

Since the FCM clustering algorithm applies normalization as defined by (5.2), this condition causes a pattern vector to have a very small amount of representation within a membership function where it should have no membership values in the ideal case. In other words, the FCM algorithm assigns a small noise as a same membership value 1/*C* to each cluster. To overcome this handicap due to the false representation, a modified α–cut method [91] is utilized to remove the noise.

Then to fit those processed membership values for each fuzzy set, a modified asymmetric Gaussian membership function defined by (5.3) is chosen for the Adaptive Membership Function Scheme (AMFS) that provides more flexibility.

$$\mu_{ij} = \mu_{ij1} \times \mu_{ij2} \tag{5.3}$$

$$\mu_{ij1} = e^{-\left(\frac{x_{kj}-v_{ij1}}{\sigma_{ij1}}\right)^2} \times Index_{v_{ij1}} + (1 - Index_{v_{ij1}})$$

$$\mu_{ij2} = e^{-\left(\frac{x_{kj}-v_{ij2}}{\sigma_{ij2}}\right)^2} \times Index_{v_{ij2}} + (1 - Index_{v_{ij2}})$$

$$if\ x_{jk} \le v_{ij1} \quad Index_{v_{ij1}} = 1,\ otherwise\ 0$$

$$x_{jk} \ge v_{ij2} \quad Index_{v_{ij2}} = 1,\ otherwise\ 0$$

The membership value μ_{ij} is determined by the j-th feature value of the k-th pattern vector, x_{kj}, the cluster prototype value for the j-th feature of the i-th cluster, v_{ij}, and two different standard deviations, σ_1 and σ_2. The Levenberg-Marquardt type non-linear least square fit is utilized to estimate the parameters, $\{v_{ij1}\ v_{ij2}\ \sigma_{ij1}\ \sigma_{ij2}\}$ for each membership function for each fuzzy cluster. The initial values of cluster prototypes v_{ij} are obtained from the final cluster prototypes using the FCM, and deviations $\sigma_{ij1}, \sigma_{ij2}$ are initialized to the average deviation of pattern vectors in each cluster. The height of this modified asymmetric Gaussian membership function initialized as 1.0, but it is able to be controlled to be less than 1.0 during the fitting process when $v_{ij1} > v_{ij2}$. This characteristic regarding the height of the membership function provides the proposed rough-fuzzy inference system with more flexibility to model the best shapes of the training data using Gaussian basis functions.

For example, the automatic generation of membership function is carried out for one feature, petal length of the Iris data set in pattern classification scheme. Figure 5.2 shows the membership values of features in three clusters after the FCM clustering for petal length. Based on these membership values, the removal of false information is done as shown in Figure 5.3. The final fitting for the processed membership values to model membership functions is shown in Figure 5.4.

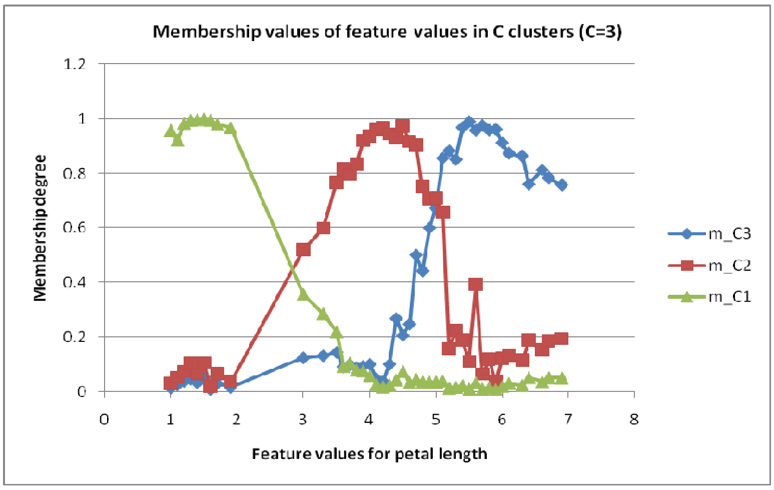


Figure 5.2 Membership values after the FCM clustering for petal length

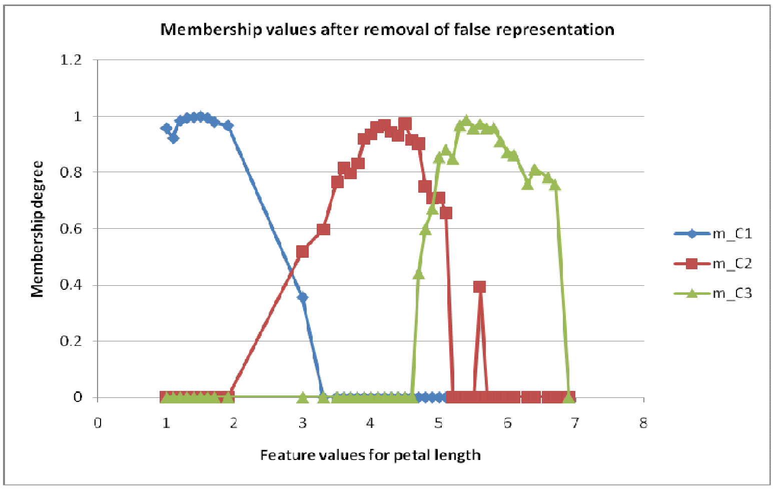


Figure 5.3 Membership values after removal of false representation

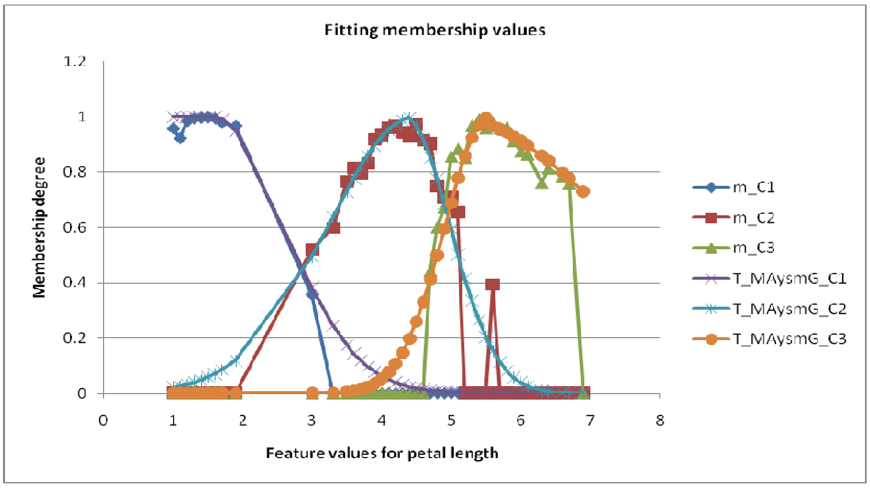


Figure 5.4 Final fitting to model membership functions for petal length

5.2 Encoding Decision Tables

Once pattern vectors of a training data set are obtained, these vectors can be considered as features that compose a conditional attribute set *C* of an information system. The associated feature used to determine its output composes a decision attribute set *D*. These feature vectors and their target vectors constitute an information system as a decision table, $DT = < U, C \cup D >$ according to the rough set theory. The original decision table from the training data may be encoded using the adaptive fuzzy partitions obtained. In the proposed system, an adaptive fuzzy partition method is applied by utilizing the Fuzzy C-Means (FCM) clustering algorithm. If the partitioned regions are described as intervals of each dimension for each feature, which replaced the numeric values of pattern vectors by the label of the fuzzy clusters, the original decision table may be converted into an encoded decision table shown as Table 5.1. For example, the encoded decision table for training data set from the Iris data is shown in Figure 5.5. The '***x#***'s in Figure 5.5 stand for arbitrarily assigned input vectors extracted from Iris data set.

Table 5.1 Encoded Decision Table using adaptive fuzzy partitions

Attributes / **Objects**	a_1	a_2	…	a_m	**DECISION**
1	3	1	…	2	1
2	2	4	…	3	2
…	…	…	…	…	…
n	1	3	…	4	3

[Encoded Decision Table for training data]

x0:	0	0	0	0	0
x1:	0	0	0	0	0
x2:	0	0	0	0	0
x3:	0	0	0	0	0
x4:	0	0	0	0	0
x5:	4	0	0	0	0
x6:	0	0	0	0	0
...					
x25:	2	0	3	3	1
x26:	2	0	3	3	1
x27:	2	0	3	3	1
x28:	4	4	4	3	1
...					
x50:	3	0	2	2	2
x51:	4	4	2	2	2
x52:	1	0	2	1	2
x53:	3	0	2	2	2
...					
x74:	2	0	2	1	2

Figure 5.5 Encoded decision table for training data set from the Iris data

5.3 Dimensionality Reduction by Rough Set Approach

The encoded decision table constructed may contain much superfluous and conflicting data. As stated earlier, one of the main advantages of the rough set methodology is that it reduces the given knowledge using the degree of dependency of attributes. This process requires finding reducts of condition attributes with respect to the decision attribute in order to obtain the smallest possible number of attributes and decision rules for higher compactness. Thus the problem associated with the number of fuzzy rules can be resolved by finding a minimal set of attributes and decision rules.

An algorithm based on the decision-relative discernibility matrix [74] with Boolean calculation is selected for the reduction of attributes. The algorithm for the reduction of attributes is as follows. Firstly, obtain the discernibility matrix, m_{ij} defined by (5.4) with respect to the decision attribute, *d*.

$$m_{ij} = \begin{cases} \{a \in C : f(x_i, a) \neq f(x_j, a)\} & \text{if} \quad d(x_i) \neq d(x_j) \\ 0 & \quad\;\; d(x_i) = d(x_j) \end{cases} \tag{5.4}$$

Then, calculate T_{ij}, the disjunctive Boolean expressions with the entries of the discernibility matrix as defined by (5.5).

$$T_{ij} = \{ \bigvee_{a_i \in m_{ij}} a_i : m_{ij} \neq 0, m_{ij} \neq \phi \} \tag{5.5}$$

Compute the Boolean expression in conjunctive normal form as defined by (5.6).

$$T = \bigwedge_{m_{ij} \neq 0, m_{ij} \neq \phi} T_{ij} \tag{5.6}$$

Calculate the Boolean expression in disjunctive normal form as defined by (5.7).

$$T^* = \bigvee_i T_i \tag{5.7}$$

Finally, find a minimal set of attributes, or a *reduct,* which has the least number of attributes from the normal form of T^*.

In addition, in order to select the *'best'* reduct amongst those candidate reducts, a fuzzy similarity measure [75] is applied. The degree of overlaps of membership functions for each feature for each reduct is measured using a fuzzy similarity measure defined by (5.8). A *reduct*, which has the smallest overlap degree on average, is chosen as the final best reduct towards better accuracy in the pattern classification scheme.

$$s(F_i, F_j) = \frac{\left(\sum_x (F_i(x) \wedge F_j(x))\right)^2}{\left(\sum_x (F_i(x) \vee F_j(x))\right)^2} \qquad (5.8)$$

A fuzzy similarity s in (5.8) is defined as the ratio of the squared sum of the degree of the intersection over the squared sum of the union between two fuzzy sets F_i, F_j. This ratio emphasizes the degree of overlaps, which is very sensitive for small changes of overlaps. As the operators of the intersection and the union calculations, the arithmetic minimum and maximum functions are used respectively.

For instance, the described attributes-reduction method is applied on the Iris data set after the FCM clustering and encoding processes were done. As a result, the obtained candidate reducts for Iris data sets are; { *Sepal Length*, *Petal Length*, *Petal Width* }, { *Sepal Width*, *Petal Length*, *Petal Width* }. The best reduct among these candidates is { *Sepal Length*, *Petal Length*, *Petal Width* }.

5.4 Generation of Decision Rules

To generate decision rules of the given information using the obtained best reduct, all training data are partitioned into corresponding disjoint equivalence classes with respect to the decision attribute. Based on these obtained equivalence classes, the decision rules are generated by applying the equation (3.36) in Chapter 3. This decision rule generation process is shown as a simple diagram in Figure 5.6. For example, the partitioned training data from the Iris data set are shown in Figure 5.7. The generated decision rules for the Iris data set are shown in Figure 5.8.

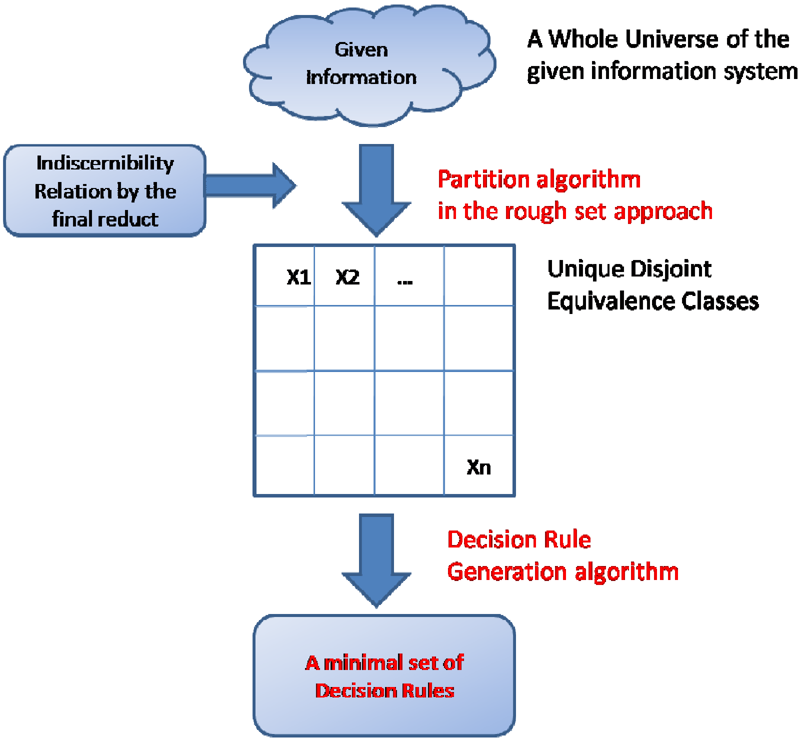


Figure 5.6 Decision Rule Generation using the final reduct, disjoint equivalence classes

[Partitioned Equivalence Classes]
X1 : x0 x1 x2 x3 x4 x6 x7 x8 x9 x11 x12 x13 x17 x19 x21 x22 x23 x24
X2 : x5 x10 x14 x15 x16 x18 x20
X3 : x25 x26 x27 x40 x49
X4 : x28 x36 x39
X5 : x29 x31 x33 x38 x43 x47 x48
X6 : x30 x41
X7 : x32 x35
X8 : x34
X9 : x37
X10 : x42 x44
X11 : x45
X12 : x46
X13 : x50 x53 x60 x66
X14 : x51 x64
X15 : x52
X16 : x54
X17 : x55 x67
X18 : x56
X19 : x57 x59 x68 x72
X20 : x58 x61 x65 x70
X21 : x62 x74
X22 : x63 x71
X23 : x69
X24 : x73

Figure 5.7 The partitioned equivalence classes for training data using the obtained best reduct

```
[The generated Decision Rules ]
R[0]: IF SL=0.00 AND PL=0.00 AND PW=0.00 THEN DECISION=0.00 : [D]
R[1]: IF SL=4.00 AND PL=0.00 AND PW=0.00 THEN DECISION=0.00 : [D]
R[2]: IF SL=2.00 AND PL=3.00 AND PW=3.00 THEN DECISION=1.00 : [D]
R[3]: IF SL=4.00 AND PL=4.00 AND PW=3.00 THEN DECISION=1.00 : [D]
R[4]: IF SL=3.00 AND PL=3.00 AND PW=3.00 THEN DECISION=1.00 : [ND]
R[5]: IF SL=4.00 AND PL=3.00 AND PW=3.00 THEN DECISION=1.00 : [D]
R[6]: IF SL=0.00 AND PL=4.00 AND PW=4.00 THEN DECISION=1.00 : [D]
R[7]: IF SL=0.00 AND PL=4.00 AND PW=3.00 THEN DECISION=1.00 : [D]
R[8]: IF SL=3.00 AND PL=4.00 AND PW=4.00 THEN DECISION=1.00 : [D]
R[9]: IF SL=4.00 AND PL=4.00 AND PW=4.00 THEN DECISION=1.00 : [D]
R[10]: IF SL=4.00 AND PL=3.00 AND PW=2.00 THEN DECISION=1.00 : [ND]
R[11]: IF SL=3.00 AND PL=4.00 AND PW=3.00 THEN DECISION=1.00 : [D]
R[12]: IF SL=3.00 AND PL=3.00 AND PW=3.00 THEN DECISION=2.00 : [ND]
R[13]: IF SL=4.00 AND PL=3.00 AND PW=2.00 THEN DECISION=2.00 : [ND]
R[14]: IF SL=3.00 AND PL=2.00 AND PW=2.00 THEN DECISION=2.00 : [D]
R[15]: IF SL=4.00 AND PL=2.00 AND PW=2.00 THEN DECISION=2.00 : [D]
R[16]: IF SL=1.00 AND PL=2.00 AND PW=1.00 THEN DECISION=2.00 : [D]
R[17]: IF SL=3.00 AND PL=2.00 AND PW=1.00 THEN DECISION=2.00 : [D]
R[18]: IF SL=1.00 AND PL=1.00 AND PW=1.00 THEN DECISION=2.00 : [D]
R[19]: IF SL=0.00 AND PL=3.00 AND PW=3.00 THEN DECISION=2.00 : [D]
R[20]: IF SL=1.00 AND PL=1.00 AND PW=2.00 THEN DECISION=2.00 : [D]
R[21]: IF SL=2.00 AND PL=2.00 AND PW=2.00 THEN DECISION=2.00 : [D]
R[22]: IF SL=2.00 AND PL=2.00 AND PW=1.00 THEN DECISION=2.00 : [D]
R[23]: IF SL=3.00 AND PL=3.00 AND PW=2.00 THEN DECISION=2.00 : [D]
```

Figure 5.8 The generated decision rules for the Iris data set

Regarding the generated decision rules, there are a couple of things to point out. Firstly, the reduct which is a minimal set of attributes reduces the number of decision rules. Accordingly, the computational complexity is less than the case when the system is using all input features to generate fuzzy rules. Secondly, it can be seen that there are Non-Deterministic (ND) rules among the generated decision rules in Figure 5.8. This means the generated rules have some conflict rules which have the same inputs for the different outputs. Obviously, these rules need to be tuned or enhanced later towards better system performance. Finally, an investigation is definitely required to examine that these rules are suitable to process the T-S type fuzzy inference. This examination will be carried out in the next section to ensure that the generated rules have a full coverage of input and output relations of the given information.

5.5 Validity Checking of Generated Decision Rules

After the generation of a minimal set of decision rules, their validity must be ascertained for use as the T-S type fuzzy inference rules. The number of antecedent fuzzy variables in the generated minimal set of decision rules may be less than the total amount of input variables of the whole fuzzy inference system. However, according to the definition of the T-S type fuzzy model [7], the T-S type fuzzy rules have a form of a combination of linear systems with all input variables as defined by Takagi and Sugeno. Hence, there is a need to investigate the validity of the generated rules in order to model the suitable T-S type fuzzy inference rules of the proposed system.

The question is whether the decision rules provide a full coverage of the information inherent within the training data. The decision rules are obtained using the final reduct, the partition algorithm, resulting equivalence classes, and decision rule generation algorithm according to the rough set approach. Since the partition algorithm divides a whole universe of the given information system into the disjoint equivalence classes via the corresponding indiscernibility relation determined by the reduct, all equivalence classes are unique in terms of their input and output relations. Using these equivalence classes, the decision rule generation algorithm produces the minimal set of decision rules. Also the partitions using the final reduct provide the same partition in the case when all attributes are deployed. In other words, the minimal set of decision rules obtained offers a full coverage of the given information and a unique set of partitions of the training data with respect to the decision attribute.

However, the antecedent variables in the decision rules may not show all input variables since some of them have been eliminated in the reduction process. In order to form a complete numeric mapping using all input and output information according to the definition of the T-S type fuzzy model, the reduced antecedent variables should be complemented in their rules. Therefore, the T-S type fuzzy rules with input information complemented may be represented as defined by (5.9). For estimating the values of coefficients of the complemented T-S type fuzzy rules, a weighted least-squares algorithm can be deployed to minimize the additional errors from the complemented information.

$$y_i = y_i^{'} + y_i^{*} = c_{i0} + c_{i1}x_{k1} + \cdots + c_{il}x_{kl} + c_{i(l+1)}x_{k(l+1)} + \cdots + c_{im}x_{km} \tag{5.9}$$

where

$y_i^{'}$: a decision rule with reduced attributes

y_i^{*}: a complementing rule

It is crucial that this investigation regarding the validity of the generated decision rules should be carried out in the process of an automatic fuzzy rule generation to provide a full input and output relation of the given knowledge for the proposed rough-fuzzy inference systems.

5.6 Construction of ARFIS

Once parameters of antecedent membership functions are found via the FCM clustering and the T-S type fuzzy rules are obtained through the rough sets approach, the proposed framework of Adaptive Rough-Fuzzy Inference Systems (ARFIS) can be constructed. The proposed framework is built as a MISO T-S type fuzzy model as mentioned earlier. All attributes are assigned as antecedent variables with the corresponding adaptive cluster information after the FCM clustering. Through the validity checking for the generated decision rules, the T-S type fuzzy inference rules are modeled using the complemented decision rules. A type of *Generalized Modus Ponens* (GMP) compositional rules is used to form fuzzy rules in the knowledge base. The algebraic minimum operator is utilized to calculate the fuzzy *T-norm* operation (‘*AND’*) between the antecedent variables. The coefficients of the consequent variables are fitted by the least squares fitting towards the corresponding target values. The Figure 5.9 shows the construction stage of the proposed system.

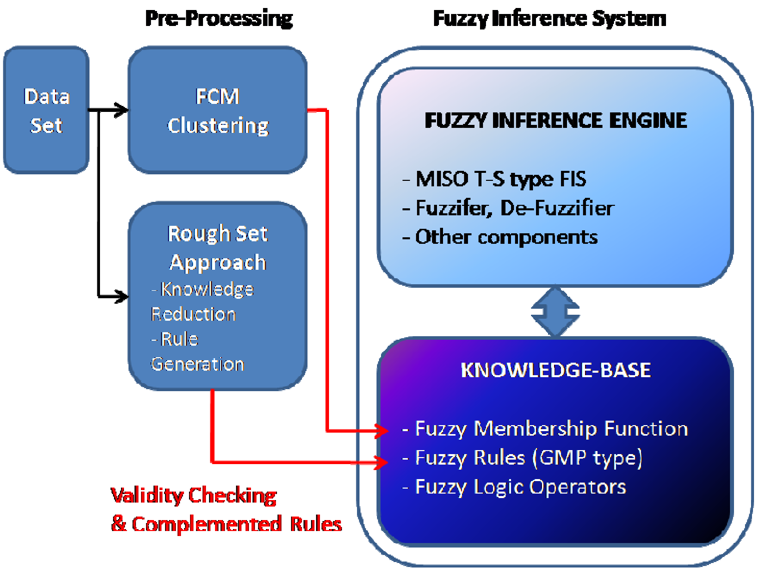


Figure 5.9 Construction stage of the proposed rough-fuzzy inference system

5.7 Adaptive Mechanism of Tuning the Knowledge-base

The performance of the proposed system needs to be evaluated and enhanced towards better achievement. After coefficients of the consequent variables of the T-S type fuzzy rules are fitted with the training data, the performance evaluation is done first with the training data to compare the RMSE measure defined by (5.10) with a user-defined error criterion. If the RMSE is not satisfactory, the adjustment of antecedent membership functions is carried out on the training data set by employing the *Polak-Ribiere* conjugate gradient algorithm to minimize the system error.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(err_i)^2}{n}}\ ,\ err_i = y^d - y^o \tag{5.10}$$

where

err_i : the error between the desired output and the actual output

n : the number of pattern vectors

y^d : the desired output

y^o : the actual output from the fuzzy inference system at one epoch

Moreover, the modeled T-S type fuzzy rules from the generated decision rules need to be tuned to get rid of some conflicts as mentioned earlier. For those rules which have same inputs but different outputs, the average rule firings can be calculated during the first system evaluation stage. The rules have the lower average rule firings can be deleted in the rule base towards higher accuracy of the proposed system. The system evaluation and the associated adjustment mechanism mentioned are shown in Figure 5.10.

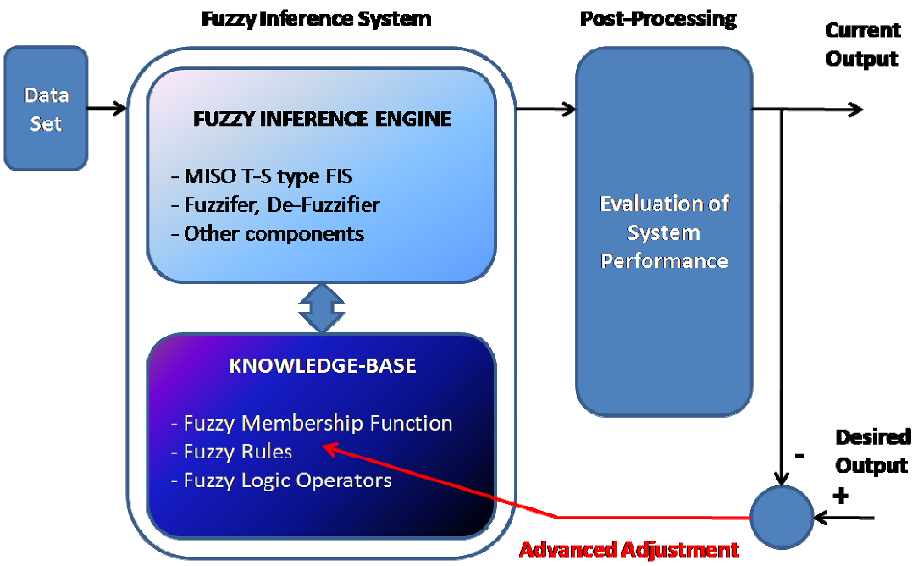


Figure 5.10 The system evaluation and the adjustment mechanism of the proposed system

5.8 Performance Metrics

Measuring how well a system performs its tasks could be dependent on the objectives of the system. In most cases, however, it is often intuitionally straightforward. Most common approaches are; 1) a calculation of the percentage of correct answers in testing environment and the subsequent comparison of it with other results under the same conditions, and 2) a measurement of the system performance regarding on robustness when the additional noise are applied. In this section, three performance metrics are described and used to measure the performance of the proposed system within the scheme of computational intelligence systems.

5.8.1 Cross Validation

Cross validation is a method that enables the system to estimate how well a system performs on the testing data which are unseen in the training phase. It is actually a prediction of generalization ability of a system. The procedure of cross validation is as follows. The whole data sets are partitioned into subsets for training and testing according to the partition strategy. The partition strategy could be simple, such as a selection of training or testing data sets in order from the first pattern vector. Or it could be a random selection in, for instance, Jackknife estimate [115].

The *N*-fold cross validation using random selection is deployed for measuring the system performance of the proposed framework of ARFIS. Using this technique, the whole data set is partitioned into *N* different subsets, and *N*-1 subsets are used for training and one subset is used for testing in an iterative procedure. This process is continued repeatedly *N* times until all *N* subsets are used for testing purposes. The testing results after *N* iterations can be calculated to produce an error estimate and the variance of the results decreases as *N* increases. The choice of *N* is dependent on the characteristics of data sets and the problem domain as well..

5.8.2 Root-Mean-Square-Error (RMSE)

The RMSE error measure is a widely-used one for the differences between target outputs from a supervised model and actual current outputs from an estimated model. In statistics, the MSE of an estimator is one of many approaches to quantify the amount that an estimator differs from the true value of the quantity being predicted. The difference occurs due to the no-account of an estimator on some information which could produce a more accurate estimation. The RMSE error measure defined by (5.10) is employed for the proposed system to obtain the system error rates.

5.8.3 Confusion Matrices

For those systems that have multiple output classes, the confusion matrix is very useful to calculate the percent correct as an analogous performance metric. If there are m classes, an $m \times m$ matrix is constructed and its rows and columns are designated as target classes and estimated classes, respectively. The value in each entry in the matrix represents the total number of pattern vectors predicted in the testing environment. The diagonal entries are the instances classified correctly. After the classification by a system, the resulting confusion matrix can be used to produce an average percent correct for the system by adding all the entries on the diagonal of the matrix and dividing the result by the number of classes. For the proposed system, the confusion matrix has been applied to applications to obtain the accuracy of the system performance.

5.9 Summary

As the main contribution of this thesis, this chapter presented a development of a new framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS) to generate membership functions and rules automatically and to resolve the existing difficulties regarding the number of fuzzy rules within the Rough-Fuzzy hybridization scheme. The subsequent adaptive mechanism is proposed in the system evaluation and enhancement stage towards higher system performance. In addition, the investigation on the generated decision rules is carried out regarding their validity for use in the T-S type fuzzy inference process to ensure that the rules have a full coverage of input and output relations of the given information. It is seen that the generated rules have the same capacity to represent the given knowledge by the aid of the rough set approach. Also it is noted that the rules are complemented with the eliminated input features to model them as the T-S type fuzzy rules. This is a crucial process to generate the T-S type fuzzy rules automatically in the construction stage of the proposed system. Some performance metrics have been applied to show how well the proposed system performs.

For the further analysis of the proposed system within the context of a framework of fuzzy inference systems, there must be some more systematic measures to show that the proposed system is a suitable tool to achieve the system objectives. More issues and systematic metrics are described and suggested later to enhance the proposed system.

Chapter 6

6. Applications

The proposed framework of ARFIS has been applied to some applications to show its viability as a framework for rough-fuzzy inference systems. The application domains are; pattern classification, face recognition, and mobile robot navigation. As the nature of problems in each application is different, the proposed framework has been tailored for each application according to its system objectives.

The rest of this chapter is organized as follows.

In section 6.1, the proposed system has been applied to pattern classification on three data sets; the Fisher's Iris, the Wisconsin Breast Cancer, and the new Wisconsin Diagnostic Breast Cancer data sets retrieved from [92]. The objective of the application is to classify those data sets towards higher accuracy using less number of input variables and fuzzy rules via the proposed approach. The results from each data set have been compared with other existing pattern classifiers. It has been shown that the results are very satisfactory and competitive.

The face recognition is carried out as the next application in section 6.2 on the face image database from the MIT Media Lab [100]. The aim is to reduce the high dimensionality of face images using the rough set approach and to recognize them using the proposed fuzzy inference system. The results of this task have been compared with other approaches to show that the results are very encouraging with much more research potential in this field of research.

The final application for the proposed system is the mobile robot navigation and is described in section 6.3. The wall-following robotic behavior has been performed for a variety of environments; straight walls, circular walls, arbitrary-shaped walls, and sharp-corner (90 degrees) walls. The methods employed for a comparison of the system performance are the following; a Bang-Bang controller, a PID controller, a conventional (standard) fuzzy controller, a GA-fuzzy controller, and a rough-fuzzy controller. The system objective is to get a better control with a reduced number of input features and fuzzy control rules by applying the rough set approach. The results have shown that the quality of the control using the proposed rough-fuzzy inference system is relatively better than other controllers with satisfactory performance.

6.1 Pattern Classification

In the past, many different approaches have been suggested to achieve a higher accuracy on a variety of data sets in the pattern classification scheme. For instance, as reported in [91], a conventional method [85] and fuzzy-based classifiers; the Adaptive Fuzzy Leader Clustering (AFLC) [86], Wu and Chen's algorithm [87], the Fuzzy Entropy-Based Fuzzy Classifier (FEBFC) [90], and the Influential Rule Search Scheme (IRSS) [91] have been applied on the Fisher's Iris data [92] and the Wisconsin Breast Cancer data set [92] to achieve better performance. However, some of these approaches still have difficulties especially with the number of fuzzy rules when a higher dimensional data set is applied, because the size of their knowledge-base in fuzzy inference systems is directly associated with the computational complexity and the system performance.

The Fisher's Iris, the Wisconsin Breast Cancer, and the new Wisconsin Diagnostic Breast Cancer data sets were obtained from the UCI machine learning repository [92] for the experiment. For each data set, the experiment was carried out under the following same conditions in the training and the testing environment. The details of the experiment for each data set are described in the next sections.

- The N-fold cross validation using random selection with $N = 10$
- The FCM clustering with $C = 5$
- The modified α-cut method with $\alpha = 0.02$ for the removal of false representation during the FCM clustering
- The initialized values of deviations of membership functions as the average deviation within the clusters obtained the FCM clustering
- The reduction of the given information with respect to the decision attribute
- The assigned user-error criterion for the RMSE measure as 0.2 in the adjustment process of antecedent membership functions
- The fine tuning of modeled fuzzy rules for the removal of conflict rules
- The different 10 independent runs for the average of results
- The statistical test ANOVA (analysis of variance) applied to demonstrate the statistical differences in results produced by different approaches

6.1.1 The Fisher's Iris Data

The Fisher's Iris data set [92] contains 150 pattern vectors with four features (Sepal Length, Sepal Width, Petal Length, and Petal Width) and one output of three classes (Iris Setosa, Iris Versicolor, and Iris Virginica). The scatter plots for the Fisher's Iris data are shown in Figure 6.1.

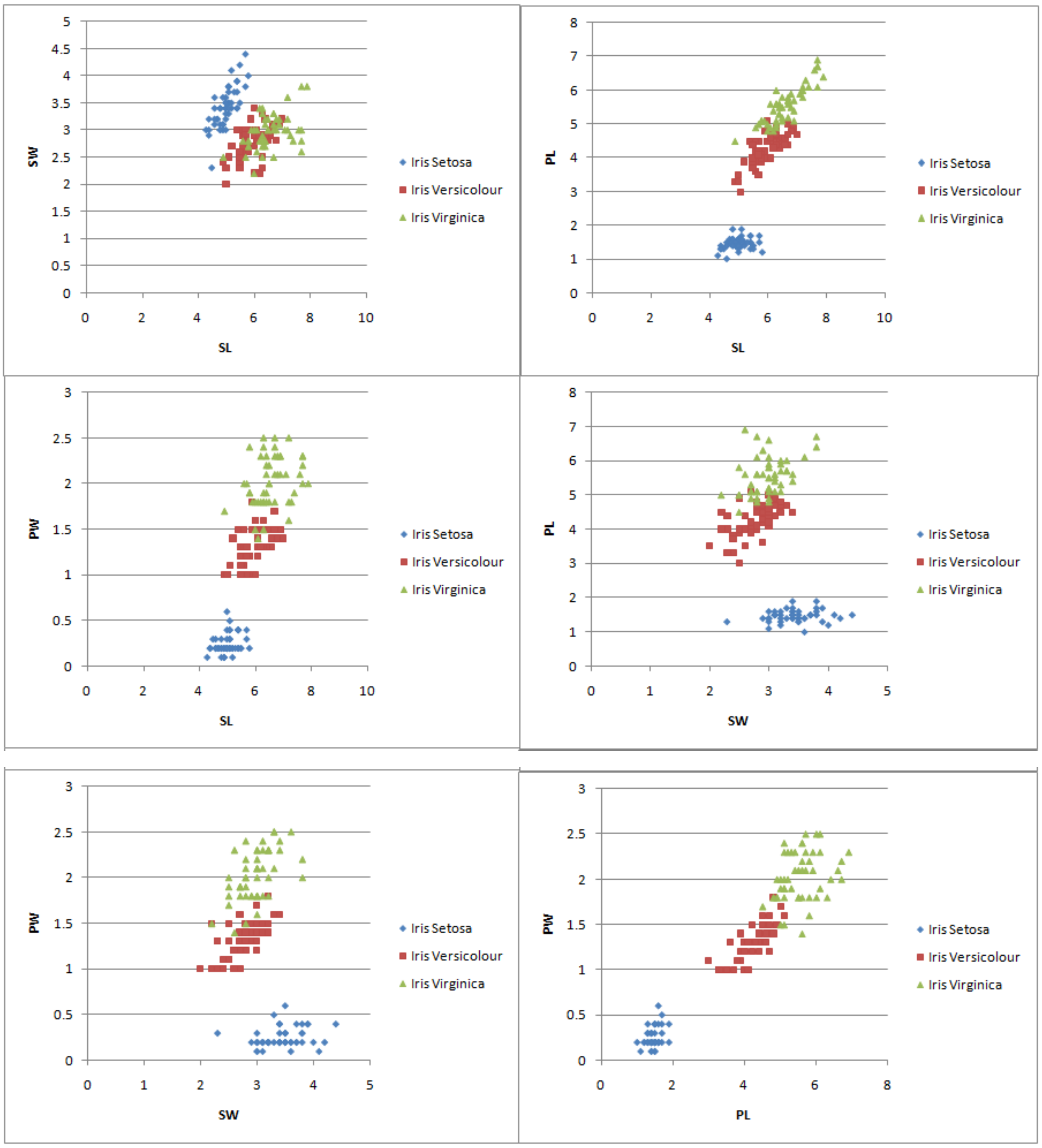


Figure 6.1 The scatter plots for the Fisher's Iris Data Set

The antecedent membership functions were automatically generated for Iris data set and one example for an attribute, 'Sepal Length' is shown in Figure 6.2. It can be seen each membership function was fitted to different shapes of asymmetric Gaussian function.

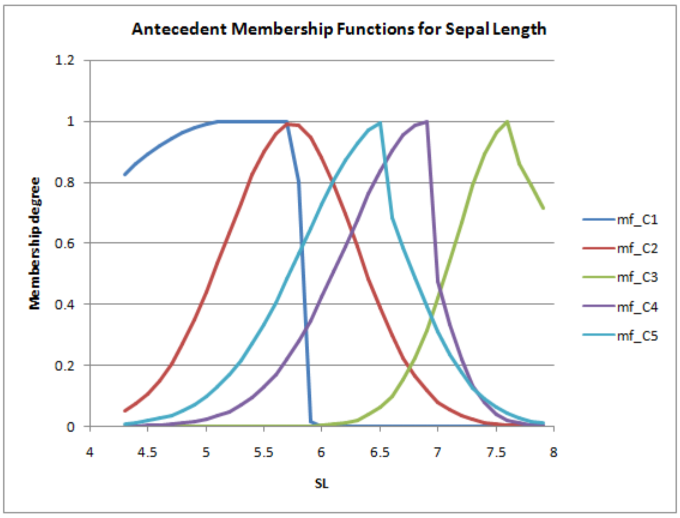


Figure 6.2 The generated antecedent membership functions for Sepal Length

Utilizing 10-fold cross validation, the proposed ARFIS randomly selected 9 subsets for training and 1 subset for testing in each fold. The validation process was repeated 10 times, with each of the data vectors used exactly once as the validation data.

The best final reduct calculated for the Iris data set was {Sepal Length, Petal Length, Petal Width} (3 features). The number of reduced fuzzy rules was 23 on average after the 10 independent runs. One set of the generated rules is already shown in Fig 5.8 as an example. In contrast to the number of the generated fuzzy rules in the IRSS [91] which increased exponentially as $5 \times 5 \times 5 \times 5$, our proposed system achieved a massive amount of reduction, 96 % for this data set, on the number of fuzzy rules. Accordingly, the computational complexity was reduced effectively by the proposed rough set approach. The result of the classification accuracy as the average percent correct is shown in Table 6.1 with results from other classifiers for a comparison of system performance. It can be seen that the classification accuracy produced by the proposed framework of ARFIS is very competitive compared to results of other classification approaches.

Table 6.1 Classification accuracy on the Fisher's Iris Data

Algorithms	Setosa (%)	Versicolor (%)	Virginica (%)	Average classification ratio (%) ($\pm\sigma$)
GVS [85]	100	94.00	94.00	96.00
AFLC [86]	100	86.00	100	95.33
Wu and Chen [87]	100	93.38	95.24	96.21
FEBFC [90]				97.12
IRSS [91]	100	92.00	96.00	96.00
ARFIS [93], [113]	100	92.05	96.67	96.24 ($\pm$2.18)

6.1.2 The Wisconsin Breast Cancer Data

The Wisconsin Breast Cancer data set [92] is used to test our proposed system on a higher dimensional data. It has 699 medical instances with nine attributes (ClumpThickness, UniformityOfCellSize, UniformityOfCellShape, MarginalAdhesion, SingleEpithelialCellSize, BareNuclei, BlandChromatin, NormalNucleoli, Mitoses) and one output of two classes (Benign and Malignant). In order to create subsets for training and testing, a couple of steps were applied as follows. Amongst all 699 pattern vectors from the original data set, samples that include missing attributes ('?') were firstly removed. Then, the 10-fold cross validation technique was employed to estimate the classification results. Scatter plots for this Cancer data set are shown in Figure 6.3.

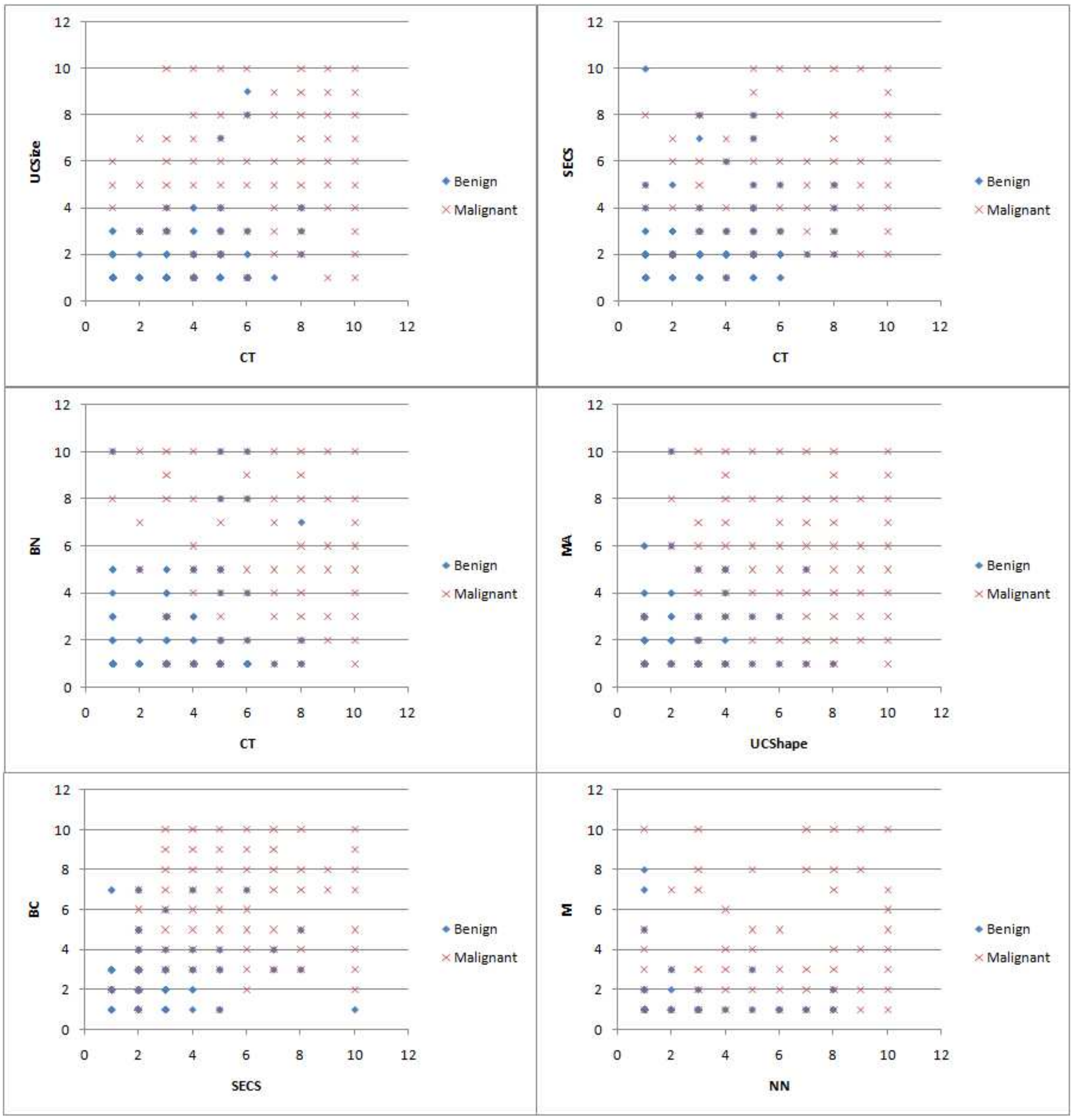


Figure 6.3 The scatter plots for the Wisconsin Breast Cancer Data Set

The experiment was performed in the same manner utilized for the Fisher's Iris data set. The generated antecedent membership functions for an attribute, Clump Thickness are shown in Figure 6.4. The best final reduct found for this Cancer data set was {CT, UCSize, UCShape, MA, SECS, NN} (6 features). The number of reduced fuzzy rules was 128 on average after the 10 simulations. As shown above, the effective knowledge-reduction is achieved on the number of attributes and decision rules on this high dimensional data set. Regarding the results of the classification accuracy, it is shown with its statistics in Table 6.2 to compare with results from other approaches. Note that our proposed ARFIS produced very competitive and much higher accuracy on a higher dimensional data set in the pattern classification scheme. This can be seen in Figure 6.5 as a difference in distributions of classification accuracy for approaches applied on this data set. Based on results in Table 6.2, the statistical test ANOVA was applied to show the statistical difference of outcomes produced by Setiono's Neuro Classifier [88] and ARFIS, and it is shown in Table 6.3.

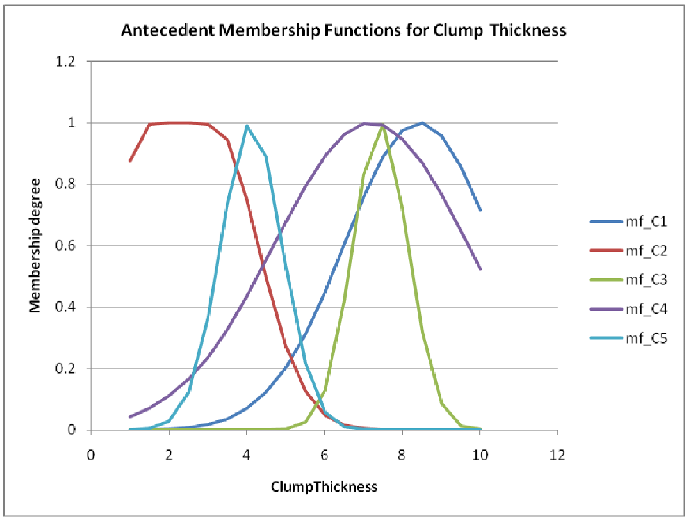


Figure 6.4 The generated antecedent membership functions for Clump Thickness

Table 6.2 Classification accuracy on Wisconsin Breast Cancer Data

Algorithms	Accuracy (%) ($\pm\sigma$)
Setiono's Neuro Classifier [88]	93.99 ($\pm$4.81)
MSC [89]	94.90
FEBFC [90]	95.14
IRSS [91]	95.89
ARFIS [93], [113]	96.47 ($\pm$2.05)

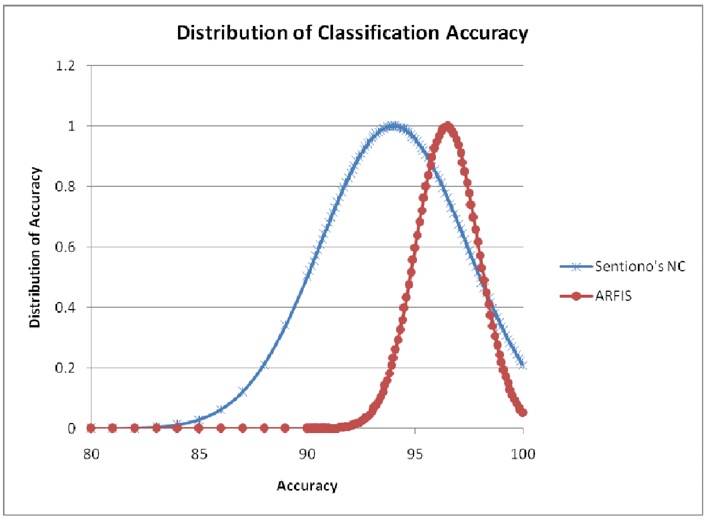


Figure 6.5 The distribution of classification accuracy for approaches on Wisconsin Breast Cancer Data

Table 6.3 The ANOVA test on classification results on WBC data

Source	Sum of Squares (SS)	Degree of freedom (df)	Mean of Squares (MS)	F	p
Between-Groups (Col)	6.085	1	6.085	76.404	0.0001
Within-Groups (Row)	14.698	110	0.133613	1.678	
Error	8.761	110	0.079		
Total	29.543	221			

In brief, the ANOVA test compares means by examining the F ratio, which is the ratio of between-groups variance divided by within-groups variance. The F ratio effectively provides an estimate of the extent to which the distributions from two (or more) groups or conditions overlap. The more the distributions overlap the less likely it is that the means differ and vice versa. The larger difference in the means causes F to be greater so increasing the likelihood of a significant difference between the means.

From Table 6.3, it can be seen from a simple visual inspection on test result that the between-groups MS (variance) is far greater than the within-groups MS. This means that the variability across the different approaches of classification is much higher than the other which is from one subject to another for points in distribution of classification accuracy. In addition, the ARFIS has achieved higher accuracy with smaller deviation than result from Setiono's. Therefore, it can be stated that the ANOVA test on classification results indicated that the proposed system achieved statistical significance of differences in means of classification accuracy on the Wisconsin Breast Cancer data set. This statistical test showed another quantitative proof that the proposed system produced better results than other approaches on this data set in the pattern classification scheme.

6.1.3 The Wisconsin Diagnostic Breast Cancer (WDBC) Data

The Wisconsin Diagnostic Breast Cancer (WDBC) data set [92] is one of the later versions of Wisconsin Breast Cancer data with different medical features computed from a digitized image of a fine needle aspirate of a breast mass. These attributes represent characteristics of the cell nuclei present in the image. It has 569 instances in total with 30 real-valued input features and the same output of two classes (Benign and Malignant). Ten real-valued features are computed for each cell nucleus; radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The mean, standard error, and worst or largest of these features were computed for each image, resulting in 30 features. The scatter plots for the Diagnostic Cancer data are shown in Figure 6.6.

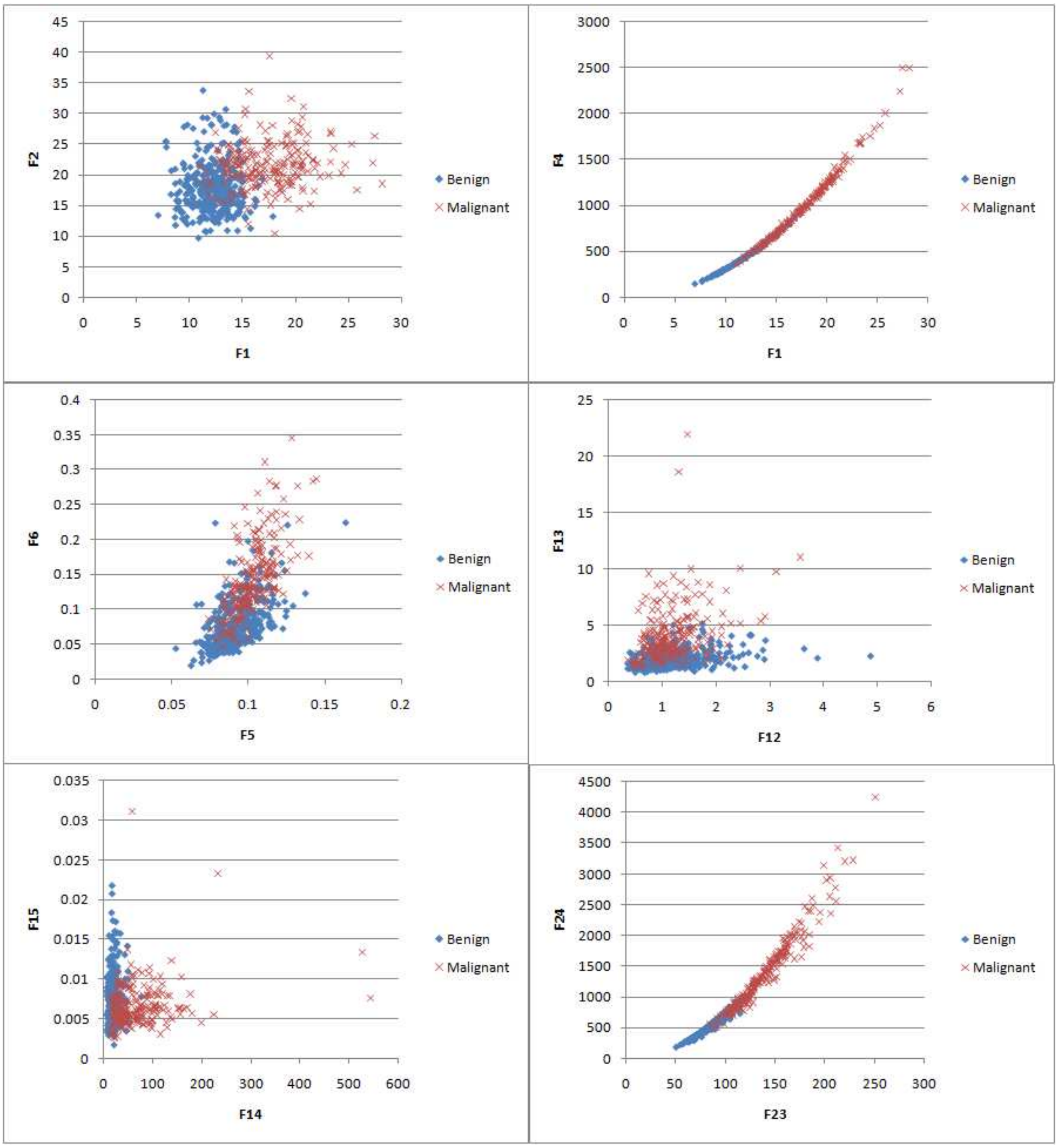


Figure 6.6 The scatter plots of the Wisconsin Diagnostic Breast Cancer Data Set

As the Wisconsin Diagnostic Breast Cancer data set has been popularly used in the literature [94], [95] for one of the benchmark of higher dimensional data, the proposed system was applied to this data set under the same procedure. The generated antecedent membership functions for the first attribute are shown in Figure 6.7. The best final reduct found for this data set was {F1, F3, F4, F5, F6, F8, F9, F12, F15, F18, F19, F21, F23, F24, F25, F26, F29} (17 features). The reduced number of fuzzy rules was 278 on average after the 10 simulations. As shown, the reduction is achieved much more effectively on the number of attributes and fuzzy rules on such a high dimensional data set. The results of the classification accuracy are shown with statistics in Table 6.4 for a comparison with results produced by other approaches. The distribution of classification accuracy for approaches is shown in Figure 6.8. Note that our proposed framework of ARFIS produced much higher accuracy compared to most of other approaches on this higher dimensional data. The statistical ANOVA test was also applied to results produced by other approaches and ARFIS, and it is shown in Table 6.5.

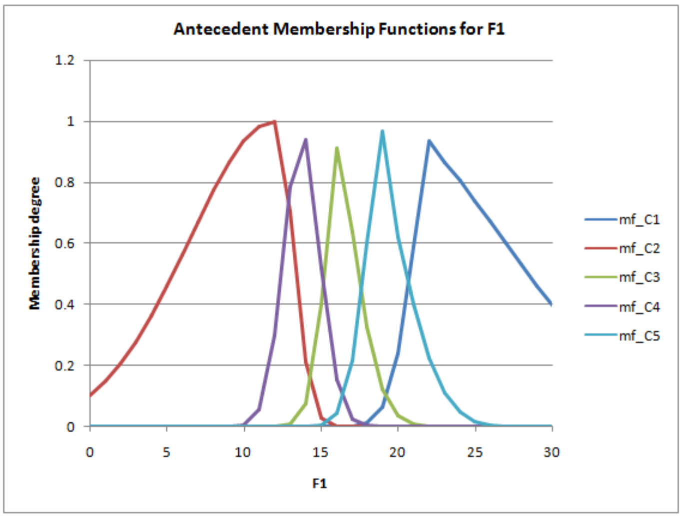


Figure 6.7 The generated antecedent membership functions for the first feature, F1

Table 6.4 Classification Accuracy on Wisconsin Diagnostic Breast Cancer Data

Algorithms	Accuracy (%) ($\pm\sigma$)
PLV [94]	93.15
RB [95]	93.69 ($\pm$3.38 max)
KD [95]	94.93 ($\pm$2.12 max)
SS1 [95]	96.11 ($\pm$0.51 max)
ARFIS	95.59 ($\pm$1.41)

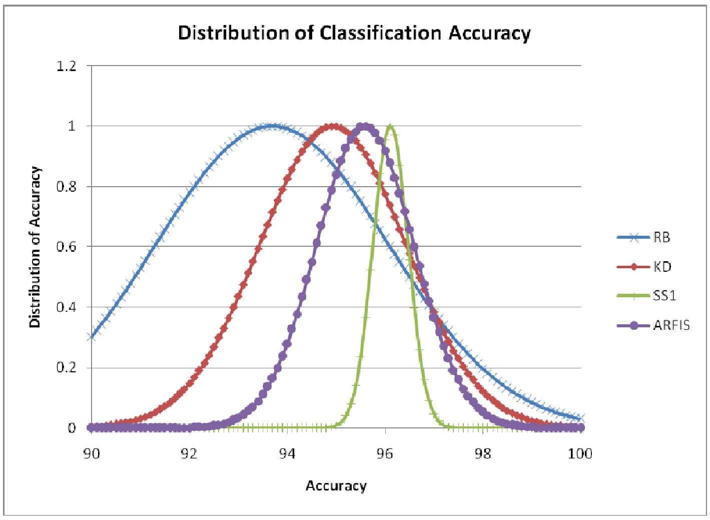


Figure 6.8 The distribution of classification accuracy for approaches on Wisconsin Diagnostic Breast Cancer Data

Table 6.5 The ANOVA test on classification results on WDBC data

Source	Sum of Squares (SS)	Degree of freedom (df)	Mean of Squares (MS)	F	p
Between-Groups (Col)	10.827	3	3.609	75.532	0.0001
Within-Groups (Row)	28.975	110	0.263	5.513	
Error	15.767	330	0.048		
Total	55.568	443			

The ANOVA test results in Table 6.5 showed that the between-groups variance which is across the different approaches is greater than that of MS for within-groups case. Also the F ratio shown is quantitatively large than another which is driven by within-groups source. This means that the proposed ARFIS produced statistical significance in differences of means on the Wisconsin Diagnostic Breast Cancer data set. As seen, even though the best result was not produced by ARFIS on this data set, the proposed ARFIS achieved a comparatively higher classification result and a statistical significance on difference in means of accuracy in pattern classification.

6.1.4 Conclusion

This chapter described the application of the proposed framework of ARFIS in the pattern classification scheme. In order to assess the viability of the proposed system, three data sets were utilized to show the performance of ARFIS using different dimensionality and complexity of the example data set. The pattern classification task for these data sets is carried out under the same conditions as stated earlier. In each section for the experiment for each data set, their characteristics were mentioned with brief overview and the results were shown with the reduced size of the input features and fuzzy rules. Regarding the classification accuracy, results on each data set were compared with other approaches in the context of the pattern classification.

By the comparison of the classification results, it can be stated that the proposed framework of ARFIS has a very efficient knowledge-reduction process to reduce the high complexity of the given information, and excellent generalization ability with the proposed adaptive mechanism to adjust its knowledge-base towards better achievement in the pattern classification scheme. Based on the system performance of ARFIS in this application, it can be seen that there is much potential for this research in pattern recognition on high dimensional data within the context of fuzzy inference systems.

6.2 Face Recognition

Recently, the face recognition has been applied to a variety of practical systems such as identification systems in airports, security surveillance systems and so on. The most well known approach is Eigenface model [96], [97] using the Principal Component Analysis (PCA). The PCA is an unsupervised statistical method that finds the most relevant information to represent the given data. It has been used widely for the past decades in the face recognition area. However, the dimensionality of features from face images is obviously too high to process them. The processed data using the PCA approach have still too much computational complexity to calculate. In order to resolve this problem, further analysis of face recognition using the ICA [98] along with the comparisons with the PCA, and the rough set approach [99] have recently been considered for the effective feature-reduction in this research area to achieve better results. The advantages of both the rough set approach and the T-S type fuzzy model are combined to develop a T-S type PCA-Rough-Fuzzy Inference System for face recognition. Also, a theoretical similarity in the representation of the given information as a combination of linear systems in between the T-S type fuzzy model and the Eigenface model has led the authors to propose a new face recognition approach using the proposed system framework.

6.2.1 Eigenface Model

An information theory approach for coding and decoding face images has led M. Turk and A. Pentland to develop a face recognition system [96], [97] using principal components of face images. In order to extract the most relevant information of face images from a database, the PCA (Principal Component Analysis) method has been applied. The eigenvalues of the covariance matrix of the set of face images have been used to rank the corresponding eigenvectors. An eigenvector associated with the largest eigenvalue holds the most relevant information that contributes to describe the distribution of face images in the '*face space*.' Using these most relevant eigenvectors considered as a set of features, to characterize the variation of face images, each face image in a training set can be represented as a form of a linear combination of the eigenvectors. Each eigenvector can be displayed as a 2-D vector image that is called an '*eigenface*' due to its face-like appearance.

The algorithm of the face recognition process using the eigenface model [96] takes the following steps.

1. Calculate the eigenfaces of the training set of face images.
2. If a face image is fed to the system as an input, calculate a set of weights based upon the *M'* eigenfaces and the input vector by projecting the input face image onto each of eigenfaces.
3. When the *Euclidean* distance between the projection vector and the face space is sufficiently small, determine the input face image as a face.
4. If the input is a face image, classify its projection vector as a known or unknown individual.

In the training set of face images $\Gamma_1, \Gamma_2, \ldots, \Gamma_M$, the average face is defined by (6.1).

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i \quad i=1, 2, \ldots, M \tag{6.1}$$

The difference between the average face and each face image can be obtained by (6.2).

$$\Phi_i = \Gamma_i - \Psi \qquad i=1, 2, \ldots, M \tag{6.2}$$

The difference images above are then used to find a set of *M* orthonormal vectors, or eigenvectors $\mathbf{u}_k$, and the associated eigenvalues λ_k of the covariance matrix defined by (6.3),

$$C = \frac{1}{M}\sum_{i}^{M}\Phi_i\Phi_i^T = AA^T \quad i=1, 2, \ldots, M \tag{6.3}$$

where

$$A = [\Phi_1 \ \Phi_2 \ \ldots \ \Phi_M]$$

If the dimension of each face image is N^2, the size of this covariance matrix is $N^2 \times N^2$, which means its computational complexity is extremely high. However, it is possible to determine the eigenvectors by solving the $M \times M$ matrix [97]. This calculation reduces the dimension from N^2 (the order of the number of pixels in the images) to M (the order of the number of images in the training set) to become computationally feasible.

After the eigenfaces are obtained as above, the face recognition becomes a pattern classification task. The M' eigenfaces which are selected based on the largest associated eigenvalues span a M'-dimensional subspace of the original N^2 image space. Now a new face image Γ is projected onto the face space by each eigenface as defined by (6.4).

$$\omega_k = u_k^T(\Gamma - \Psi) \quad k=1, 2, \ldots, M' \tag{6.4}$$

The M' number of projection vectors ω_k form a vector, or a '*face class*' Ω, as defined by (6.5). A *face class* describes the contribution of each eigenface to the representation of the input face image, treating eigenfaces as a basis set of face images.

$$\Omega^T = [\omega_1 \omega_2 \cdots \omega_{M'}] \tag{6.5}$$

The *Euclidean* distance, ϵ_k is used in (6.6) to find the nearest face class k that provides the best representation of an input face image.

$$\epsilon_k = \|(\Omega - \Omega_k)\| \quad \Omega_k\text{: } k\text{-th face class vector} \tag{6.6}$$

Finally, a face image is classified to class k when a minimum ϵ_k is below a threshold θ_ϵ, otherwise the face is classified as unknown.

6.2.2 Design of a PCA-Rough-Fuzzy Inference System

The proposed PCA Rough-Fuzzy Inference System [113] is built as a MISO T-S type fuzzy system. All attributes (projection weights, ω_k) from the training face images are set as the antecedent fuzzy variables with equally distributed fuzzy clusters. Using these fuzzy partitions, the given information is converted as an encoded decision table. The best final reduct is found via the proposed knowledge-reduction process. The decision rules were generated using the best reduct and the rule generation algorithm in the rough set approach. Through the validity investigation for the generated decision rules, the minimal set of decision rules are used as a set of T-S type fuzzy inference rules. The coefficients of the T-S type fuzzy rules are estimated with target values which are distances from the origin point to the projection of the input vector in an *M'*-dimensional subspace. During the estimation process, a system performance evaluation is carried out using the RMSE measure towards higher accuracy. The functional block diagram of the construction of the proposed system is shown in Figure 6.9.

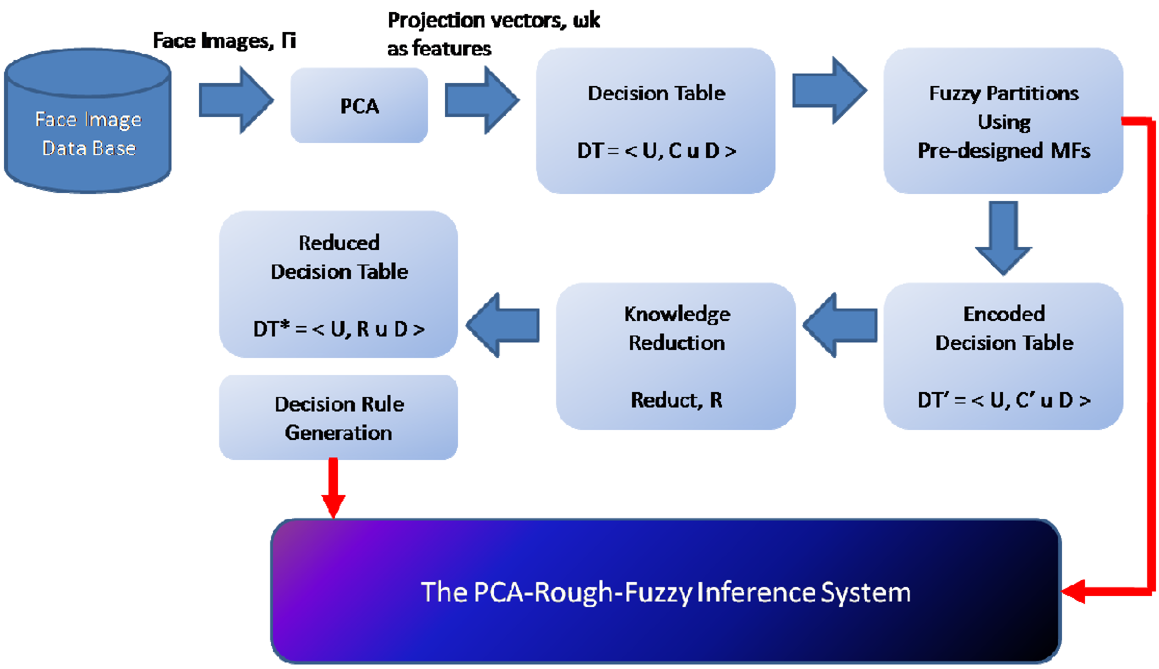


Figure 6.9 The functional block diagram of the proposed PCA-Rough-Fuzzy Inference System

In the recognition stage of the proposed system, a set of testing face images is projected onto the pre-determined eigenvectors through the PCA method. If a projection vector of an input face image is near the face space, then the input is classified as a *face*. If not, it is a *non-face*. The corresponding projection vectors are normalized for each feature and they are used to build a decision table. This decision table is then filtered and reduced by the best reduct found in the construction phase of the proposed system. The feature values from the reduced decision table are fed into the constructed T-S type fuzzy inference system to identify their personal IDs. An arbitrary threshold is applied on the difference errors between the desired value and the T-S type fuzzy output to classify an input face image into one of the person IDs.

6.2.3 Results

Since the previous successful work [93] on the pattern classification task, the face recognition scheme is applied as another application to the proposed framework of Adaptive Rough-Fuzzy Inference Systems (ARFIS) in [113]. In order to assess the viability of the proposed system, the MIT Media Labs face image database [100] was employed. In the MIT face image database, each face of 16 people was digitized 27 times, varying the head orientation, the lighting, and the scale in three types for each. The images were then filtered and sub-sampled to produce six levels of a Gaussian pyramid. Example raw face images from the MIT Media Labs face image database are shown in Figure 6.10.

Figure 6.10 Example face images from the MIT Media Labs face image data base

In our experiment, only full-scale and frontal face images for 16 people were considered for assessing the performance of the proposed system. For each individual, 5 face images were used in the training set and 4 face images were used in the testing set. Once the system was implemented and tested with those data sets, the training and the testing data were swapped using the *N*-fold cross validation technique.

In the construction stage, the pre-designed antecedent membership functions were used for fuzzy partitions. Regarding the knowledge-reduction process, the number of input features was reduced from 15 to 8 and the number of the generated rules was reduced to

16 on average after the 10 independent runs. Based on the reduced knowledge, the proposed system was built as a MISO T-S type fuzzy inference system.

In the recognition phase, the face images in the training set were classified as one of the sixteen individuals and no faces were rejected as unknown. The recognition accuracy was **93.75%** on average after 10 independent runs. During the performance evaluation stage after the estimation of the coefficients of the T-S type fuzzy system, the RMSE measure between the actual and the target outputs was satisfactory. Towards better performance, an adjustment process on the antecedent membership functions and fuzzy rules can be applied using the conjugate gradient algorithm based on the RMSE measure. If this adaptive process were to be employed, it is expected that the system performance on this task would be much better with the flexibility of the knowledge-base of the proposed system.

Given this potential of the proposed system, the results in the experiment can be seen as encouraging and satisfactory when it is compared with the results of the conventional Eigenface model reported in [96], [98]. Note that the proposed system effectively reduced the high computational complexity of the given problem by deploying the rough set approach. Also the employed T-S type fuzzy inference system achieved the classification task using its good generalization ability.

There are a couple of key points identified in the experiment. These are as the followings.

1. The recognition accuracy using reduced feature sets, or reducts, is higher than the results using the full feature set.
2. The most relevant features associated with the largest eigenvalues are in reduced feature sets at all times.

The first point is ensured by the rough set theory, but no literature to date has clarified the characteristic for the second one in terms of the theoretical linkage between the PCA and the rough set theory. This fact may indicate positive potential for the rough set approach on the knowledge-reduction scheme in the pattern recognition area.

6.2.4 Conclusion

This section described the face recognition task as another application of the proposed framework of ARFIS. In order to test the proposed system, the face images from the MIT Media Labs face database were employed to perform the given task. The experiment is carried out to construct the proposed PCA-Rough-Fuzzy Inference System from the training set and to test the system with the testing set under the system configurations as mentioned. In the experiment, the properties of the chosen face images were described and the results were shown with the reduced number of the input features and the fuzzy rules. The recognition accuracy on the face images was compared with the result of the conventional Eigenface model.

Based on the comparison of the system performance, it can be said that the system performance of the proposed PCA-Rough-Fuzzy Inference System on the face recognition scheme produced encouraging and satisfactory results with a reduced number of input features and fuzzy rules. This achievement was done by deploying the effective knowledge-reduction process of the rough set approach and by employing the excellent generalization work of the T-S type fuzzy model. It is expected that the recognition accuracy would be much better if the advanced tuning process was applied to the knowledge-base of the proposed system towards better system performance. Note that the proposed system has future potential in this field of research in terms of the theoretical development or clarification of the relations in the knowledge-reduction process in between the PCA and the rough set theory.

6.3 Mobile Robot Navigation

In the field of mobile robotics, most of the robotic behaviors have been hampered by a large number of uncertainties in real world environments. The uncertainty factors mainly come from the sensory inputs, which are often very noisy and sometimes unreliable. Fuzzy logic has been applied to handle this problem, which is widely deployed to achieve goals in robotic control [8], [9], [101]. However, due to the lack of a learning property of conventional fuzzy systems, there has been a need to tune their system parameters of the knowledge-base. For the past few years, many researchers have suggested various methods of learning for fuzzy systems towards better adaptability to the external environment [12], [13], [14], [15], [80], [102].

In this chapter, we propose a new system for a Rough-Fuzzy Controller (RFC) for robotic behavior, a wall-following navigation. The controller uses rough-membership functions to improve its uncertainty reasoning. This ‘rough-fuzziness’ [103], [104] allows the whole system to analyze its environments in a more robust and reliable manner. A database from the conventional fuzzy system on the input and output feature domain has been generated as *a priori* knowledge for the proposed system. Using this set of sample data, the partition process on the given data has been carried out to produce their equivalence classes according to the rough set theory. While the robot follows the given wall, the rough-fuzzy membership degree of an input vector has been calculated to create the ‘rough-fuzziness’ of the input to perform better analysis on the uncertainty of the environment.

The proposed system has been tested in a number of environments with the *EyeSim* simulator [105], [106] and a real robot, *LabBot* [106]. Experimental results have shown that the best system performance has been carried out using the proposed rough-fuzzy system for the wall-following behavior as compared with other controllers including a Bang-Bang controller, a PID controller, a conventional (standard) fuzzy controller, and an adaptive fuzzy controller using GA (Genetic Algorithms).

6.3.1 Rough-Fuzzy Membership Functions

The rough-fuzzy membership function [103], [104] of an input $x \in U$ for a given fuzzy set F can be defined by (6.7).

$$r_{apr_R(F)}(x) = \frac{cardinality([x]_R \cap F)}{cardinality([x]_R)} \tag{6.7}$$

where

cardinality(F): the cardinality of the given fuzzy set

$$cardinality(F) \equiv \sum_{x \in U} \mu_F(x) \tag{6.8}$$

$\mu_F(x)$: a conventional fuzzy membership function of an input x for a fuzzy set F

In the literature [104], the input pattern vectors that have the same input representation of a pattern x_i form a parallelepiped space on the space of input patterns. This parallelepiped contains all the patterns from the equivalence class $[x_i]$. The 'roughness' is created in the parallelepiped when the parallelepiped contains more than one pattern and when these patterns have different fuzzy membership values. Also, the 'fuzziness' appears in the parallelepiped when the fuzzy membership values lie in (0, 1). Therefore, the presence of both the roughness and the fuzziness creates the 'rough–fuzziness'. In terms of the spatial structure, the rough–fuzzy membership of the pattern x_i is the volume occupied by the overlapped space divided by the volume of the complete parallelepiped. In other words, the volume of the overlapped space is approximated by the weighted number of patterns in the space, where the weight of each pattern is quantified by its fuzzy membership value. The concept of the rough-fuzzy membership function mentioned above is shown in Figure 6.11.

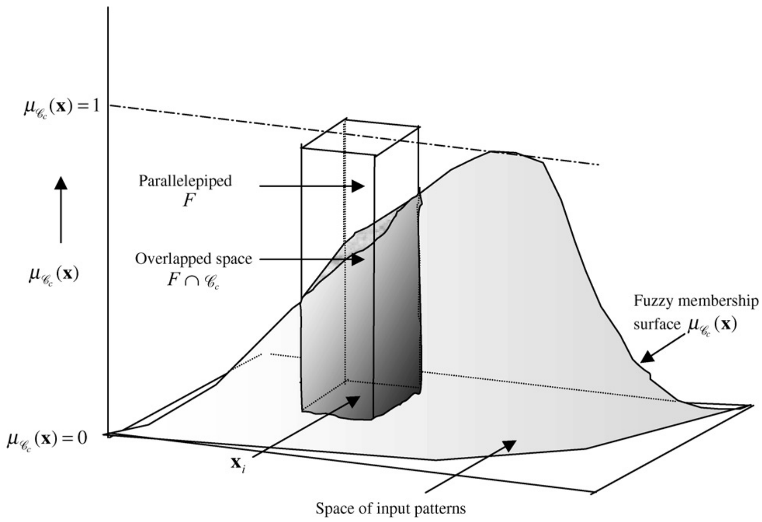


Figure 6.11 The concept of the rough-fuzzy membership functions. Adapted from the literature [104]

6.3.2 Design of a Rough-Fuzzy Controller

The proposed Rough-Fuzzy Controller (RFC) is built on the basis of a conventional fuzzy system and the rough set approach is utilized to analyze the uncertainty of the fuzziness of the input data. The rough-fuzzy membership functions were designed on top of the pre-designed fuzzy system to enhance the uncertainty reasoning. The wall following robotic behavior is applied to test our proposed system. It is a popular robotic task for exploring in structured or unknown environments. The quality of a good wall following behavior can be characterized by the following three conditions as discussed in [107]; to maintain a desired distance from the wall, to move at a constant velocity as high as possible, and to avoid sharp changes of direction and speed. It is assumed in our experiment that the robot follows the given wall on its left side. Following another side of the given wall can be easily implemented by changing the sensory inputs.

The design of the conventional fuzzy controller for wall following task is described first as follows. The knowledge-base is designed by the human expert in mobile robotics or a fuzzy expert with sufficient experience for the goal-specific design of a fuzzy system. In order to design the system, the choice of fuzzy logic operators is determined for use based on the characteristics of the given task and purpose. The goal in our case is to make the robot follow a given wall at the desired distance from the wall maintaining a constant maximum velocity. The fuzzy operators are selected as shown in Table 6.6 for the control scheme in mobile robot navigation.

Table 6.6 The choice of fuzzy logic operators

Operator	Method
T-norm	Algebraic min
T-conorm	Algebraic max
Implication	Mamdani
Aggregation	Algebraic max
Defuzzification	Center Of Area

The type of membership functions also has to be determined based on the characteristics of the goal. Towards better outcomes of the fuzzy system in the control scheme of mobile robot navigation, the standard Gaussian membership function is selected for sensory inputs as an antecedent variable and for heading angle as a consequent variable. The membership functions for the sensory input and the heading angle are shown in Figure 6.12 and 6.13, respectively. The sensory input is designed as a single left side sensor to measure the distance to the wall and a front one is used to detect the obstacles in front of the robot.

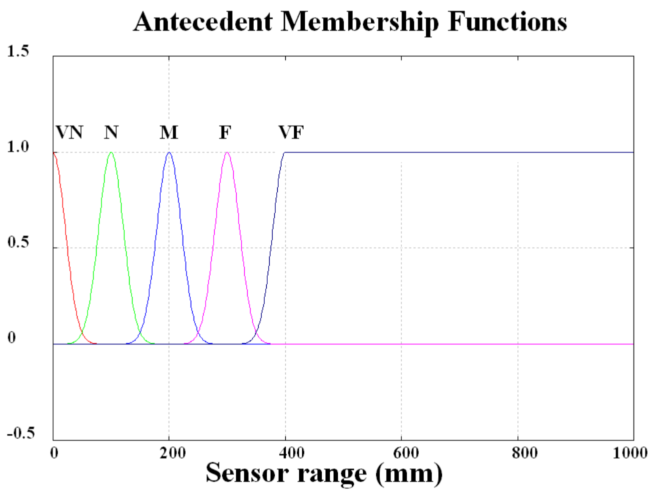


Figure 6.12 Antecedent membership functions for a sensory input

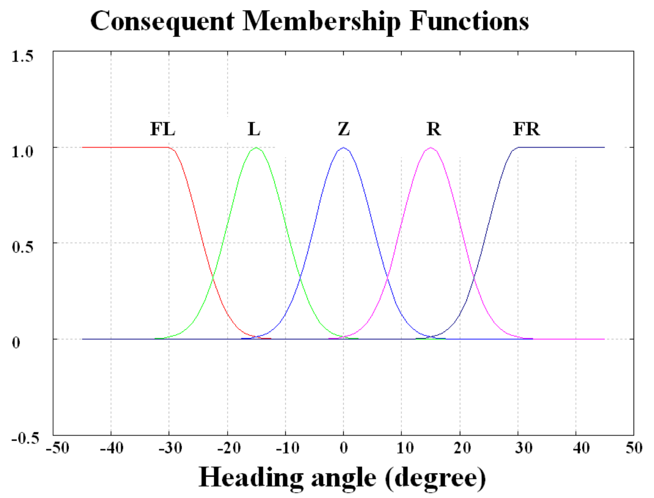


Figure 6.13 Consequent membership functions for the heading angle

The antecedent membership functions have five fuzzy linguistic variables; {VERY_NEAR (VN), NEAR (N), MODERATE (M), FAR (F), VERY_FAR (VF)}. The value of the desired distance from a given wall is 200mm and is set to the mean of the fuzzy membership function of 'MODERATE'. The other membership functions are equally partitioned within the value of two times of the desired distance, while the total universe of the discourse of the antecedent is [0, 5000] mm.

The consequent membership functions are designed using five standard Gaussian functions in the range of [-45, 45] degrees in local coordinates of the robot. They have the following fuzzy linguistic variables; {FAR_LEFT (FL), LEFT (L), ZERO (Z), RIGHT (R), FAR_RIGHT (FR)}. The mean value of 'ZERO' is set to 0.0 as the desired heading angle is perpendicular to the given wall while the robot follows it. The other membership functions are spread by equal intervals of 15 degrees.

The simple rule base of the conventional fuzzy system is constructed as Mamdani-type fuzzy inference rules and they are composed as shown in Figure 6.14. Note that the given wall is located at the left side of the robot.

R1: IF sensor_input is VERY_NEAR,	**THEN heading_angle is FAR_RIGHT.**
R2: IF sensor_input is NEAR,	**THEN heading_angle is RIGHT.**
R3: IF sensor_input is MODERATE,	**THEN heading_angle is ZERO.**
R4: IF sensor_input is FAR,	**THEN heading_angle is LEFT.**
R5: IF sensor_input is VERY_FAR,	**THEN heading_angle is FAR_LEFT.**

Figure 6.14 The designed rule base for wall following behavior

The proposed Rough-Fuzzy Controller is implemented on the basis of the designed conventional fuzzy system. The rough-fuzzy system has the ability of approximation of an input value for a given fuzzy set according to its definition. The degree of 'rough-fuzziness' of an input leads to a better representation of uncertainties linked to the environment. Using the definition of a rough-fuzzy membership function in references [103] and [104], the rough-fuzzy membership value of a pattern x_i is the volume overlapped space by an intersection between the volume of a parallelepiped for the equivalence class $[x_i]_R$ and the conventional fuzzy membership function for a given fuzzy set F as in Figure 6.15. In other words, the volume of the overlapped region is approximated by the fuzzy-weighted number of data patterns in the space of the equivalence class $[x_i]_R$.

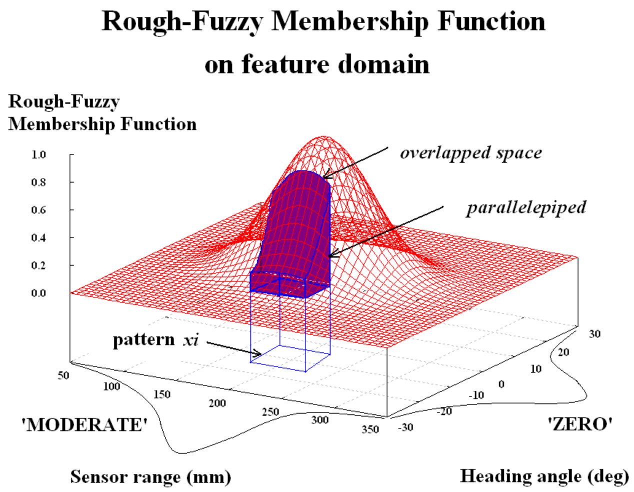


Figure 6.15 The rough-fuzzy membership function on the feature domain

In order to implement the proposed rough-fuzzy controller, a set of 1,205 input-output examples is gathered via the conventional fuzzy controller to enhance uncertainty reasoning about unknown inputs in real-time. Each data vector in the database consists of three features; a distance measured to the given wall, a linear velocity measured of a robot, and a heading angle measured while following the wall. These features are considered to calculate the rough-fuzzy membership degree of the input vector. After the partitioning process applied to the set of examples, an input vector can be categorized into one of the equivalence classes $[x]_R$ by calculating the nearest distance to the nearest equivalence class in the input space. Then the rough-fuzzy membership value of the input vector can be obtained by the cardinality of the corresponding equivalence class and the standard fuzzy membership values of features; a distance measured and a heading angle measured. When calculating the rough-fuzziness of each feature for each input vector, two given membership functions are considered. The 'MODERATE' antecedent fuzzy membership function for the sensory input and the 'ZERO' consequent membership function for the heading angle are used to produce the rough-fuzzy membership degree of the input vector. Based on the rough-fuzzy membership values of the input vector, the corresponding fuzzy inference rule is fired only when those rough-fuzzy membership values are higher than the threshold values. The threshold values are determined in a heuristic way as 0.8 for both 'MODERATE' and 'ZERO' membership degrees. The algorithm described above is summarized in Figure 6.16.

1) **Generate a set of examples for *a priori* knowledge.**
2) **Partition the set of examples gathered into equivalence classes.**
3) **While the robot follows a given wall,**
   a) Categorize the input vector into one of the partitioned equivalence classes.
   b) Calculate the rough-fuzzy membership degree of an input vector for each feature by the corresponding given fuzzy sets.
   c) Apply a threshold value to the rough-fuzzy membership to fire the corresponding fuzzy rule.

Figure 6.16 The algorithm to construct the proposed rough-fuzzy controller

6.3.3 Experiments

The mobile robot navigation scheme is applied to the proposed framework of Adaptive Rough-Fuzzy Inference Systems (ARFIS) as another application. In order to test the proposed system, the 3D mobile robot simulator, *EyeSim* [105] and a real robot, *LabBot* [106] are utilized. The experiments are carried out in five different environments to enable the robot to face different situations in navigation. The *EyeSim* 3D mobile robot simulator and the *LabBot* are shown in Figure 6.17.

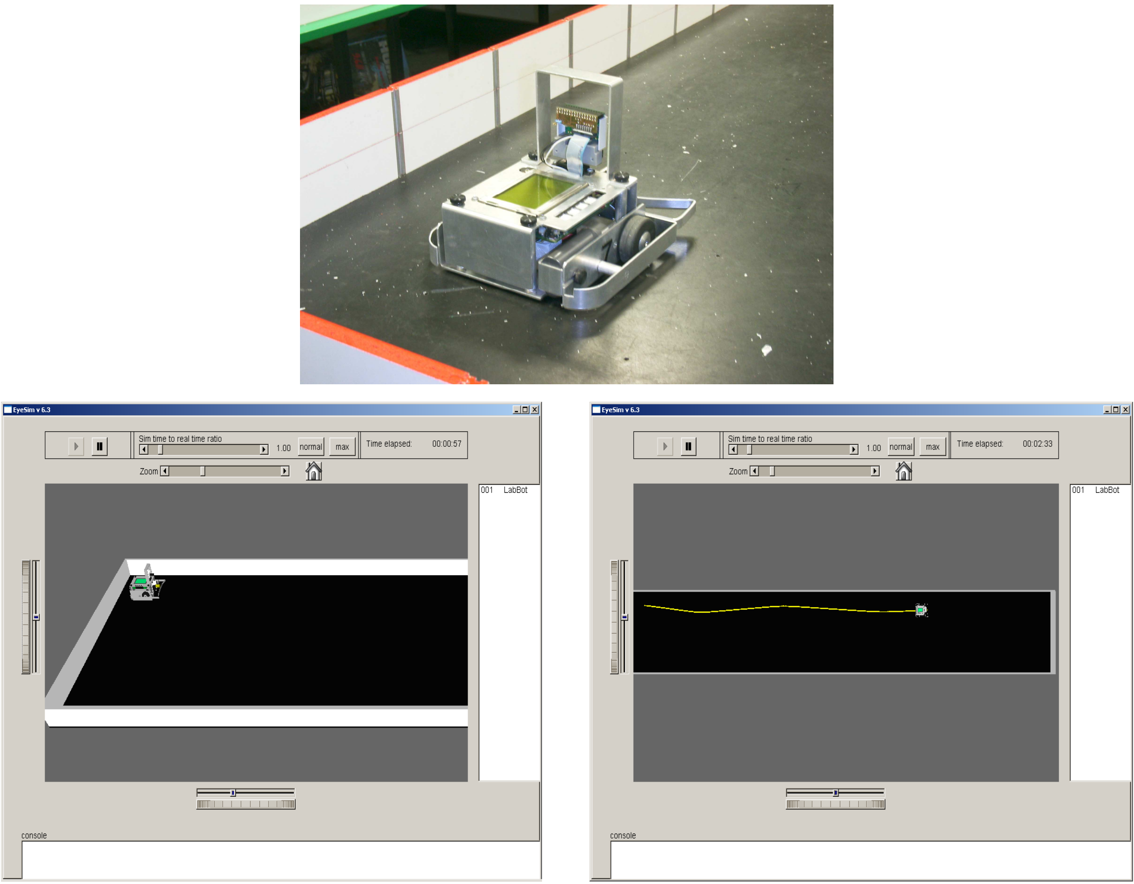


Figure 6.17 The LabBot and the EyeSim mobile robot simulator

In experiments, the desired distance from the left wall was set to 200 mm and the desired linear velocity was set to 200 mm/sec. The linear velocity of the robot was initialized as a constant value, but while the robot encounters obstacles in front 450 mm in front of the robot the velocity is designed to reduce by 10 mm/sec in each control step t_i to avoid obstacles. At the same time, the robot turns at the pre-defined angle of 30 degrees to avoid obstacles in front. As soon as the robot avoids the obstacles successfully, the linear velocity is designed to be increased to reset it up to the desired speed. This velocity control procedure and the obstacle avoidance routine are applied to all the following controllers.

For the comparison of the system performance of different methods, the following five controllers are implemented and applied for the wall-following behavior; a Bang-Bang controller, a PID controller, a conventional standard fuzzy controller, an adaptive fuzzy controller using GA, and the proposed rough-fuzzy controller.

1) The Bang-Bang control

This is the simplest control approach for the wall-following task. The Bang-Bang controller implemented here is designed to move the robot between the minimum and the maximum distances from the given wall. The maximum and the minimum distance from the wall are designed to DEISRED $\pm$ (MARGIN/2) mm, respectively. In here, the value for MARGIN is set to 90 mm. If the robot measures the minimum distance to the given wall, the robot turns away from the wall at a pre-defined heading angle of 15 degrees. In a similar way, if the robot goes too far by exceeding the maximum distance from the wall, the robot moves back toward the wall. This simplest case produces a Bang-Bang motion between the minimum and the maximum distance offsets from the wall.

2) The PID control

The conventional PID (Proportional-Integral-Derivative) controller has been widely used for control functions in the industrial process. For the past couple of decades, a number of approaches have been proposed in order to tune PID gains. For instance, the classical techniques such as Ziegler-Nichols [108], Kalman [109], and trial and error methods have been employed to adjust gains of PID controllers. Recently, more complex approaches have been utilized to re-adjust PID parameters, which are the hybrid self-organizing fuzzy PID controller [6] and an adaptive hierarchical tuning scheme for fuzzy PID controllers [4]. The classical PID controller is designed here to determine the heading angle of the robot for wall-following behavior. The mathematical formula of PID control is defined by (6.9).

$$K_p * e(t) + K_I * \int_0^t e(t)dt + K_D * \frac{de(t)}{dt} \tag{6.9}$$

where

$e(t)$: the difference between the desired and the actual heading angle

K_P, K_I, K_D : the PID gain parameters

The error function $e(t)$ is designed as the difference between the desired heading angle and the actual heading angle measured. The trial and error method is applied in order to adjust PID parameters (K_P, K_I, K_D) for the PID controller. The three gain parameters (K_P, K_I, K_D) are tuned as (0.2, 0.05, 0.3) respectively using this tuning approach to obtain a smooth motion for the robot to follow the given wall.

3) The conventional fuzzy control

The design of the conventional fuzzy system is described in the previous section.

4) The adaptive fuzzy control using GA

The classical GA approach was applied to adjust the antecedent and the consequent membership functions of the conventional fuzzy system. In this experiment, the chromosome encodes the values of the mean and the deviation of each fuzzy membership function. The heights of the Gaussian membership functions are fixed to 1.0 for simplicity. The initial seeding is given randomly with equal probability within the universe of the discourse of each parameter. The fitness function to minimize at each control step, t_i is defined by (6.10).

$$f(t_i)=\sqrt{\left\{\omega_d\cdot(d_m(t_i)-d_{des})^2+\omega_\phi\cdot(\phi_m(t_i)-\phi_{des})^2+\omega_v\cdot(v_m(t_i)-v_{des})^2\right\}} \quad (6.10)$$

where

w_d, w_φ, w_v : the weights for the distance from the wall, the heading angle of a robot, and the linear velocity, respectively

d_{des}, φ_{des}, v_{des} : the constants for the desired distance, the desired heading angle, and the desired constant linear velocity , respectively

d_m, φ_m, v_m : the variables measured at each control step, t_i

The weights, w_d, w_φ, w_v are set to 0.8, 0.1, and 0.9, respectively. The off-line learning was chosen for the tuning scheme. The parameters of antecedent membership functions were adjusted when the consequent was fixed. In a similar way, the parameters of the consequent variable were tuned while the antecedent was fixed. The final tuned antecedent and consequent membership functions are shown in Figure 6.18 and 6.19, and their final parameters are listed in Table 6.7 and 6.8.

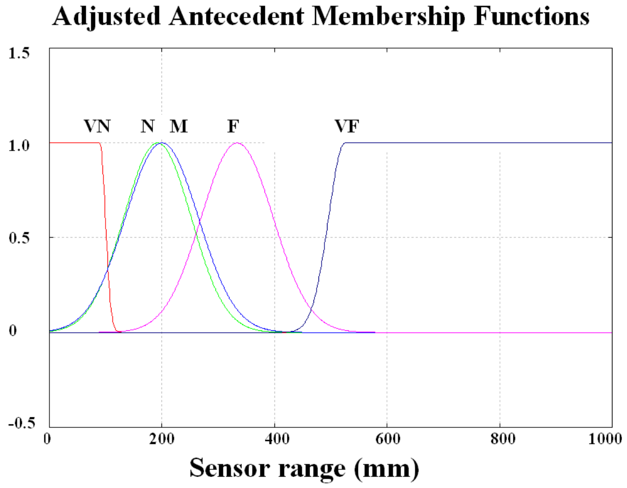


Figure 6.18 The adjusted antecedent membership functions

Table 6.7 The adjusted parameters of antecedent membership functions using GA

Tuning	Before		After	
	m	*σ*	*m*	*σ*
VN	0.0	60.0	9.62	16.06
N	100.0	60.0	115.18	88.22
M	200.0	60.0	200.0	91.22
F	300.0	60.0	338.24	89.93
VF	400.0	60.0	532.18	41.46

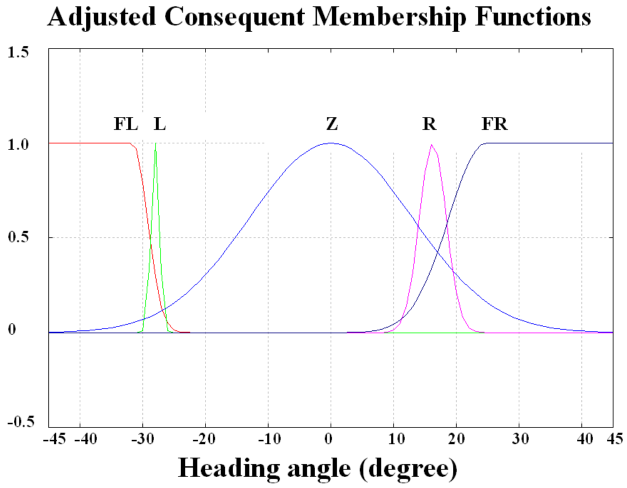


Figure 6.19 The adjusted consequent membership functions

Table 6.8 The adjusted parameters of consequent membership functions using GA

Tuning	Before		After	
	m	*σ*	*m*	*σ*
FL	-30.0	9.0	-31.56	3.21
L	-15.0	9.0	-29.18	3.64
Z	0.0	9.0	0.0	12.96
R	15.0	9.0	22.68	3.02
FR	30.0	9.0	34.83	8.29

5) The Rough-Fuzzy control

As stated earlier, the proposed rough-fuzzy controller is implemented based on the pre-designed conventional fuzzy controller. The system design of the rough-fuzzy controller and its algorithm to calculate the rough-fuzziness of an input is described in the previous section.

6.3.4 Results

The application of the proposed system on a wall-following behavior was carried out in different types of environments; straight, circular, arbitrary-shaped, and 90-degree walls. The desired distance from the left wall was set to 200 mm and the constant linear velocity was initialized to 100 mm/sec. The independent ten runs were carried out for each environment and for each control methods using a zero-error model on *EyeSim*. Also, five runs of the actual experiments were done with a real mobile robot, *LabBot.* Some of the experimental results have been published in [110], [113] and extended in [111].

In order to measure the system performance, the Performance Index (PI) is defined by (611). The formula used for calculating the PI is adapted from the fitness function defined by (6.10) excluding the component for the heading angle. Since the heading angle changes all the time, the PI value will be accumulating in response to the shape of the environments. Note that the PI values close to zero represent a better performance of the system.

$$PI(t_i) = \sqrt{\left\{\omega_d \cdot (d_m(t_i) - d_{des})^2 + \omega_v \cdot (v_m(t_i) - v_{des})^2\right\}} \qquad (6.11)$$

Regarding the test results for each control method in each environment, there are five performance metrics for wall following behavior; the average distance (mm) measured by the sensor to the left wall, the average total time (s) spent by the robot along the path, the average total length (m) of the path in environment, the average performance index (PI), and the average linear velocity (m/s) of the robot. The measurements in each table for each result are done using *EyeSim* and they are the average after 10 independent runs for each control scheme.

1) Straight wall environment

The proposed system was applied to the straight wall following task. The five control methods showed their movements for following the given wall. This experiment in straight walls was carried out using *EyeSim* first, and then done with the real robot, *LabBot*. The trajectory of each controller on *EyeSim* for a part of the straight wall following is shown in Figure 6.20 and their results are shown in Table 6.9. Based on the results shown in the figure and the table, the proposed rough-fuzzy controller performed relatively better in the straight wall environment in terms of the uncertainty reasoning and the PI measure.

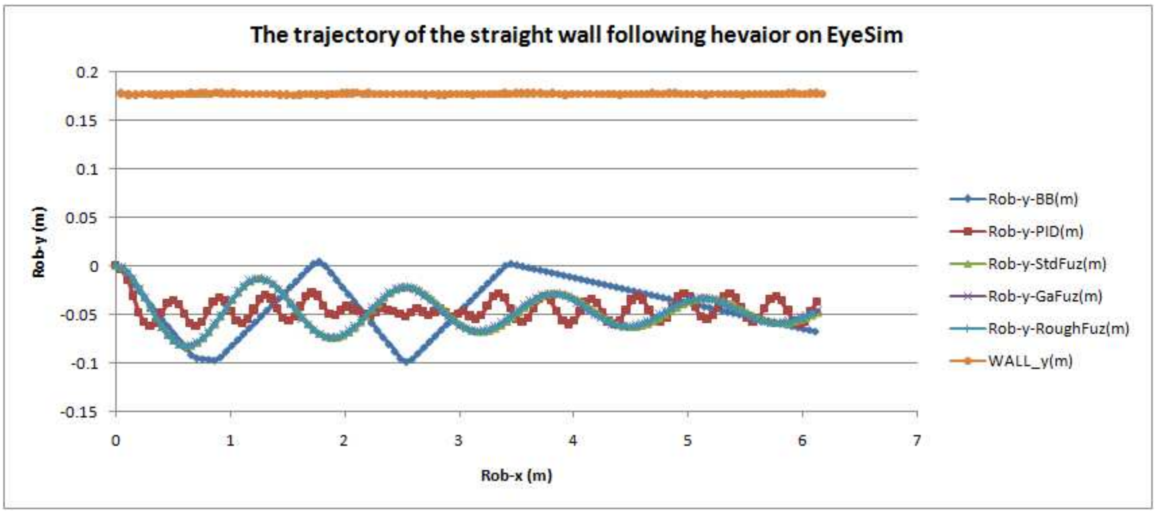


Figure 6.20 The trajectory of the straight wall following behavior on *EyeSim*

Table 6.9 The results of the straight wall following

	Dist (mm)	Times (s)	Path (m)	PI	Vel (m/s)
BB	519.77	225	56.1	129.02	0.2
PID	512.65	211	45.9	114.85	0.2
StdFuz	493.57	185	41.4	97.00	0.2
GAFuz	498.23	183	40.1	76.35	0.2
RoughFuz	503.37	185	40.5	17.92	0.2

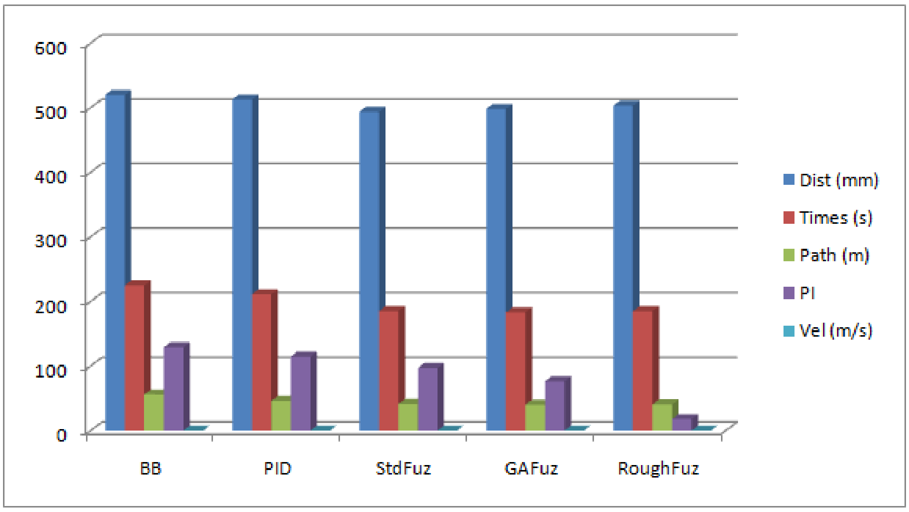
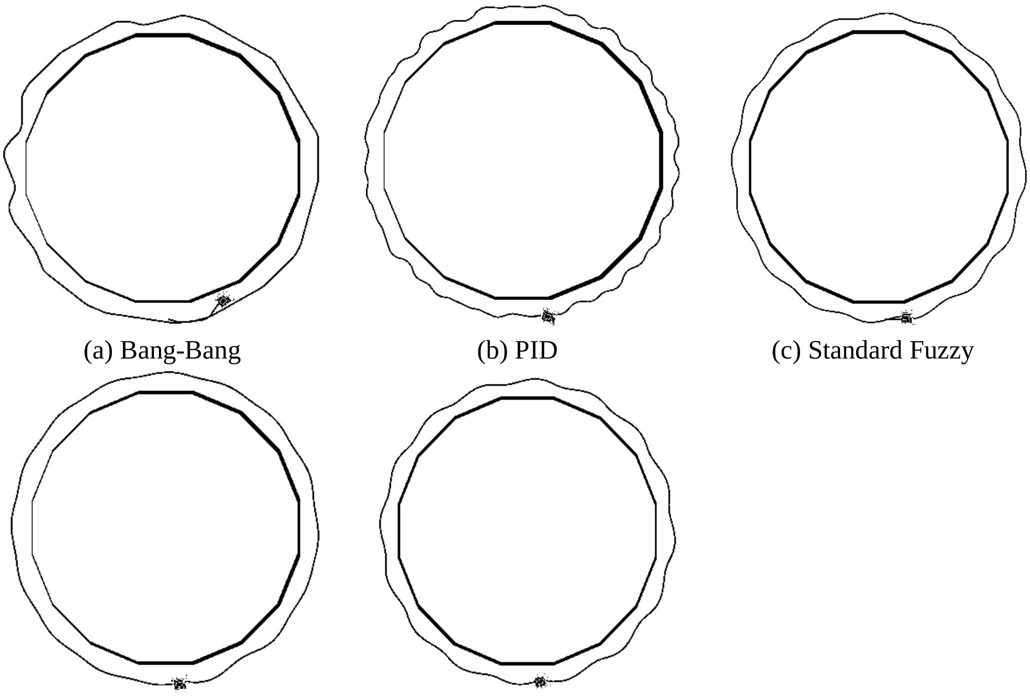


Figure 6.21 The column chart for results of the straight wall following

2) Circular wall environment

The wall following behavior was tested in a circular wall environment. The trajectories of the robot for the circular wall following for each control method are shown in Figure 6.22. The trajectories in Figure 6.22 and the results in Table 6.10 also indicated that fuzzy-based systems performed better than a Bang-Bang controller, a PID controller and also the Rough-Fuzzy controller had comparatively better outcomes over the standard fuzzy system and the GA-applied fuzzy system. Note that in this circular environment there are some oscillations of the navigation due to the corners from the line segments used to approximate the circular course.



(a) Bang-Bang (b) PID (c) Standard Fuzzy

(d) GA-Fuzzy (e) Rough-Fuzzy

Figure 6.22 The trajectory of the circular wall following behavior on *EyeSim*

Table 6.10 The results of the circular wall following

	Dist (mm)	Times (s)	Path (m)	PI	Vel (m/s)
BB	221.32	148.7	12.80	34.50	0.198
PID	214.53	134.0	12.78	29.17	0.194
StdFuz	212.38	128.5	12.66	23.92	0.2
GAFuz	210.34	131.0	12.59	26.25	0.2
RoughFuz	210.41	129.3	12.55	24.27	0.2

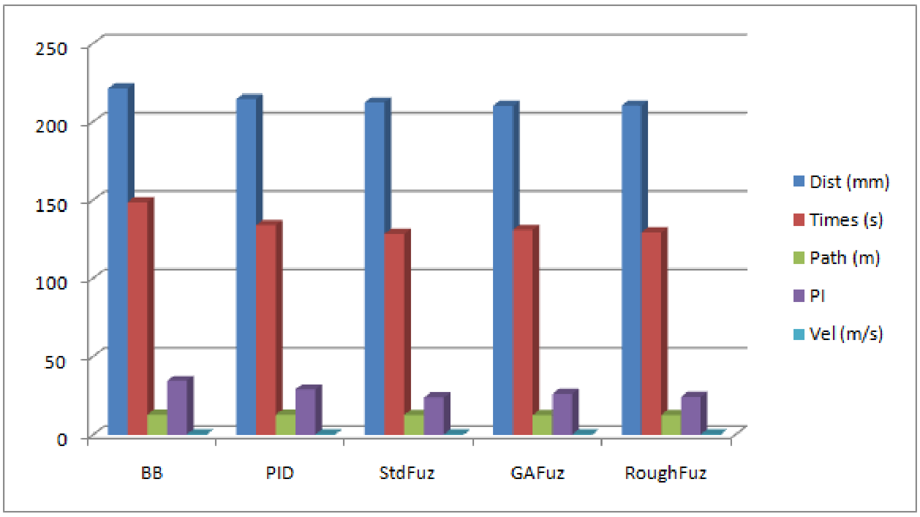
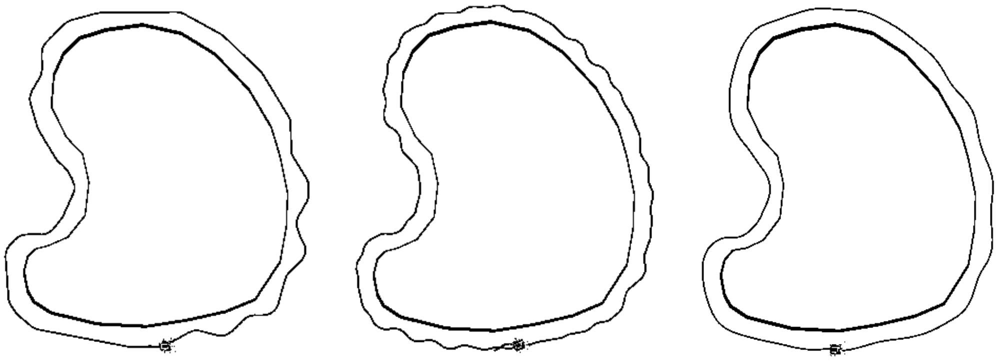


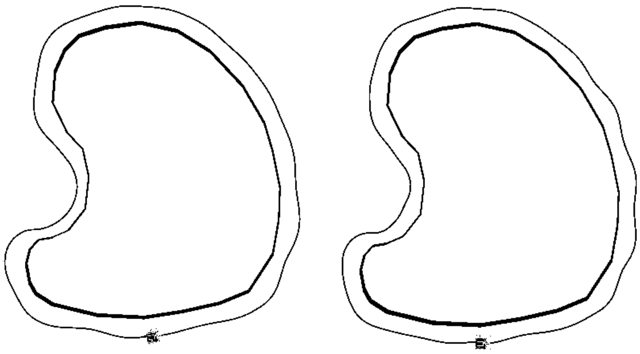
Figure 6.23 The column chart for results of the circular wall following

3) Arbitrary-shaped wall environment

All different control methods were applied to the arbitrary-shaped environment as shown in Figure 6.24. This type of environments presents local unknown arbitrary corner models for the robot to face more complicated situations during the navigation. The arbitrary-shaped environments are suitable for testing the quality of the proposed approach for wall-following task. The results of this experiment are shown in Table 6.11. The proposed Rough-Fuzzy method showed generally better results than any other method.



(a) Bang-Bang (b) PID (c) Standard Fuzzy



(d) GA-Fuzzy (e) Rough-Fuzzy

Figure 6.24 The trajectory of the arbitrary-shaped wall following behavior on *EyeSim*

Table 6.11 The results of the arbitrary-shaped wall following

	Dist (mm)	Times (s)	Path (m)	PI	Vel (m/s)
BB	230.35	172.75	13.83	51.77	0.189
PID	213.67	162.67	13.61	19.57	0.192
StdFuz	209.91	141.25	13.44	18.40	0.197
GAFuz	210.52	141.33	13.45	18.42	0.197
RoughFuz	209.83	142.33	13.43	18.38	0.197

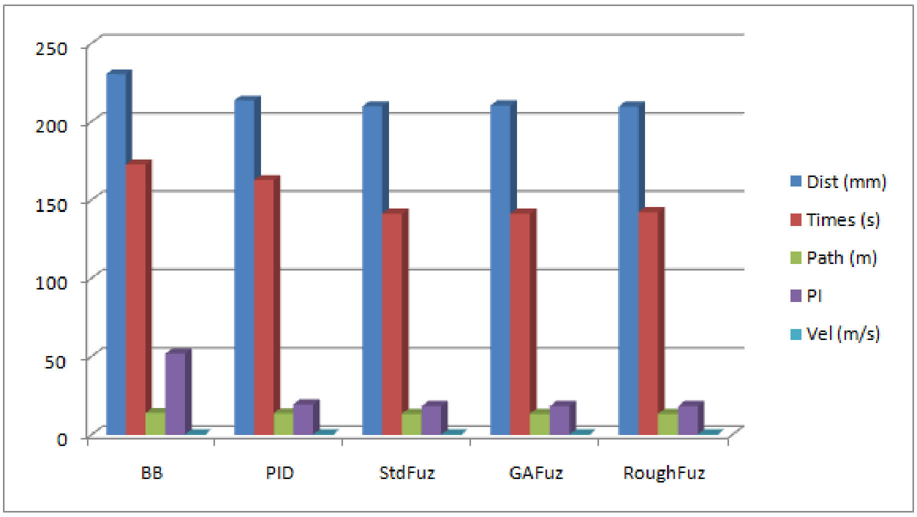
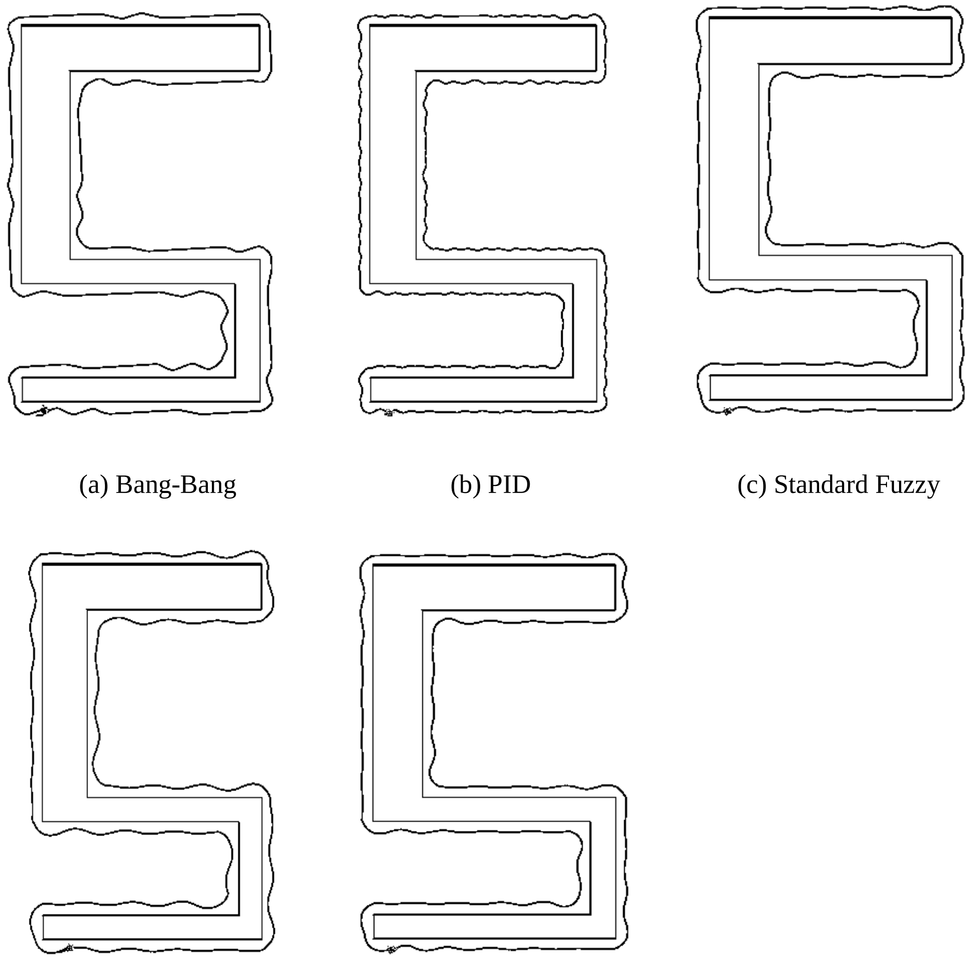


Figure 6.25 The column chart for results of the arbitrary-shaped wall following

4) Sharp-corner 90-degrees wall environment

The more complex environment including sharp 90 degrees corners shown in Figure 6.26 was used in order to compare the quality of the five controllers in sharp movements. This 90-degree wall environment has four concave and eight convex corners with a length of 39 meters. Convex corners are obviously difficult situations, because the robot sensors may not be able to detect the wall correctly at a certain control step t_i when driving at a corner. The results in Table 6.12 indicated that the proposed rough-fuzzy controller approach produced better outcomes when compared with other methods in spite of the difficulties as mentioned above.



(a) Bang-Bang (b) PID (c) Standard Fuzzy

(d) GA-Fuzzy (e) Rough-Fuzzy

Figure 6.26 The trajectory of the sharp-corners wall following behavior on *EyeSim*

Table 6.12 The results of the sharp corners wall following

	Dist (mm)	Times (s)	Path (m)	PI	Vel (m/s)
BB	246.08	575.5	54.57	70.91	0.188
PID	241.38	546.0	46.86	40.01	0.191
StdFuz	224.04	520.75	44.70	35.73	0.194
GAFuz	222.70	523.33	44.33	33.70	0.194
RoughFuz	223.49	518.0	43.11	34.35	0.194

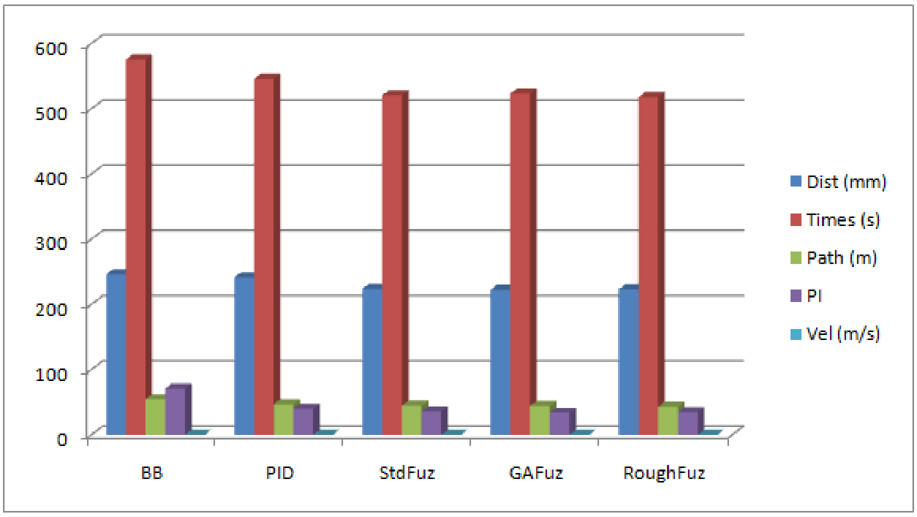


Figure 6.27 The column chart for results of the sharp corners wall following

The results in figures and tables showed that the Bang-Bang controller had the worst performance and fuzzy logic controllers outperform the Bang-Bang and the PID controllers. The adaptive fuzzy controller using GA produces better results compared with the conventional standard fuzzy controller. Our proposed rough-fuzzy controller produced better results than the standard fuzzy system. Based on the results of wall following behavior in a variety of environments, it can be stated that the proposed Rough-Fuzzy controller produced comparatively better system performance compared with other control approaches. It is expected that if the proposed system had some more multiple inputs and adaptive mechanism to tune itself, then much better uncertainty reasoning process would be done via the proposed rough-fuzzy approach.

It is crucial to remark that the improvement of the uncertainty reasoning process of the standard fuzzy system was achieved by the analysis of the “rough-fuzziness.” The rough-fuzzy approximation of an input vector led to better uncertainty reasoning process related to the environment in control scheme for mobile robot navigation.

6.3.5 Conclusion

This section mentioned the mobile robot navigation as one of the applications of the proposed framework of ARFIS. In order to apply the proposed system to the robot navigation, the input and output data samples were collected via the pre-designed conventional fuzzy system to perform the given task. The experiment is carried out to design the proposed Rough-Fuzzy Controller (RFC) from the database and to compare its performance with other control approaches. In each experiment, the wall following behavior is carried out using different five control methods in each environment.

Based on the comparison of the system performance, it can be stated that the system performance of the proposed rough-fuzzy controller in robot navigation scheme produced satisfactory and competitive results with small values of the PI (Performance Index). It was achieved by utilizing the uncertainty analysis of the rough-fuzziness of the given input. It is expected that the control output would be much enhanced if the advanced adaptive process was applied to the proposed system towards better system performance. A further study is continuing on the development of the MISO T-S type RFC on point-to-point navigation towards more robust, faster, and efficient mobile robot navigation.

Chapter 7

7. Conclusion and Discussion

In conclusion, the proposed framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS) has been developed to resolve the difficulties of the existing fuzzy systems;

1) the curse of the high dimensionality of the knowledge base when more inputs are involved,

2) the automatic generation of membership functions and rules with the absolute minimal information of the given knowledge, and

3) the adaptive mechanism for systematic tuning towards better system performance.

In order to assess the capability of the proposed system, it has been applied to a variety of applications; pattern classification on the Fisher's Iris and Wisconsin Breast Cancer data sets, face recognition on the MIT Media Lab face database, and mobile robot navigation on wall-following and point-to-point robotic behaviors. Results from the experiments have shown that the performance of the proposed ARFIS is satisfactory, and competitive. It can be said that the proposed framework is a suitable tool to achieve the given task within the context of the rough-fuzzy hybridization scheme.

In this research, there are some important issues to consider for further investigation and development of the proposed system.

Firstly, as the objective of knowledge-reduction is to obtain the absolute minimal form of the given knowledge, a comparison with other reduction methods should be investigated and carried out. For instance, the PCA, the ICA, and other reduction approaches in rough set theory should be compared on the same data set under the same constraints of the variables in their test conditions. This, of course, does not always provide the best one as a fixed case, because knowledge-reduction really depends on the data set applied, the properties of the reduction methods, and the goal of the application. However, a comparison with other knowledge-reduction approaches with statistical analysis of their results would help to provide a deeper insight to help choose the appropriate knowledge-reduction method for a particular application.

Secondly, within the T-S type fuzzy system framework, it has not yet been clarified how to exactly assign the appropriate meaningful linguistic terms to the associated membership functions at the system design and adjustment stages. This is one of the general problems of fuzzy system design for fuzzy experts. Even though, a T-S type fuzzy system incorporates very reliable system design in a particular field with a huge amount of data, this system might not work as expected for unknown inputs. After the initial construction of ARFIS, the proposed system would have a system performance evaluation stage towards better achievement including the tuning process. After the tuning stage, the initially assigned physical meaning of the membership functions would be quite different from the first ones. Then in this case the system needs a verification and/or validation for the unknown input and the expected output. Therefore, an update process of system modeling would be required to amend the initial system design of T-S type fuzzy models.

Moreover, when the proposed system handles extremely high dimensional data the proposed system needs a faster process to reduce the given knowledge, because the rough set approach calculates for all the given input features by comparing all the pairs of them. Thus, a fast knowledge-reduction method would be in great demand in an adaptive speed-up manner for higher dimensional data sets. For instance, a feature transformation into a lower dimensional data set would help to analyze and reduce the given heavy data sets. This issue will lead to many new hybrid techniques to enhance the system performance towards fast speed calculation.

Finally, as the Rough-Fuzzy hybridization has been deployed as a new trend in decision-making over the past decades, a more general mathematical model is required for the theoretical combination of fuzzy sets and rough sets. This generic model would then contribute to this field of study for researchers to model a variety of application systems with ease.

7.1 Future Work

Even though fuzzy inference systems have been used very successfully in real world applications for the past decades, they also need to have system improvement towards better performance. Due to the lack of an adaptation process within themselves, some techniques from soft computing and computational intelligence have been applied to contribute to the adaptive mechanism of the fuzzy inference systems. In general, the adjustment of fuzzy systems means the tuning of membership functions and fuzzy rules.

There are some issues to examine and problems of fuzzy systems to resolve for system enhancement. In this section, six systematic measures are described for future work to improve the proposed framework of ARFIS by investigating problems and suggesting fuzzy system metrics towards better system behavior.

Regarding the adaptation of fuzzy systems, the common problem of the tuning of membership functions is that the shape of membership functions is changed drastically so that some of fuzzy subsets lose their originally designed physical meanings. Also by the effect of the adjustment process, the fuzzy subsets can no longer cover the whole range of the input domain. In this case, the fuzzy partitioning of the input is incomplete. In other words, the fuzzy system produces no output when the value of the input is in the uncovered range. Thus, the examination on the completeness of fuzzy systems is crucial especially for the automatically generated fuzzy systems from data.

Prior to further discussion, it is required to define the term, "**completeness**". As Jin suggested in [76], a fuzzy system is said to be complete if

1) the fuzzy partitioning for each input is complete and
2) the rule structure of fuzzy rules is complete.

The fuzzy system is incomplete if one of these conditions is violated.

In this section, the first condition is described further. For the second condition, it is explained in the later section for compactness of fuzzy systems from the rough set perspective.

After the adjustment of the membership functions, the fuzzy partitions of input variables are no longer complete, because the adaptation produces re-distribution of membership functions to minimize or maximize the designed objective function to optimize the fuzzy system. However, there is an issue of 'over-fitting' in the optimization process. The over-fitting of membership functions causes the problems of incompleteness of fuzzy partitions, the loss of the physical meanings of them, and the resulting lack of distinguishability (or interpretability) of fuzzy subsets.

In order to avoid the over-fitting of membership functions, some systematic procedures or measures must be considered in the optimization process. For the system enhancement of ARFIS, a fuzzy similarity measure is used to check the completeness of fuzzy partitions of input variables and to preserve the distinguishability of them. The fuzzy similarity measure defined by the equation (5.8) in chapter 5 is employed. The

Figure 7.1 shows the concept of the similarity of two neighboring fuzzy sets F_i, F_j using the fuzzy similarity measure.

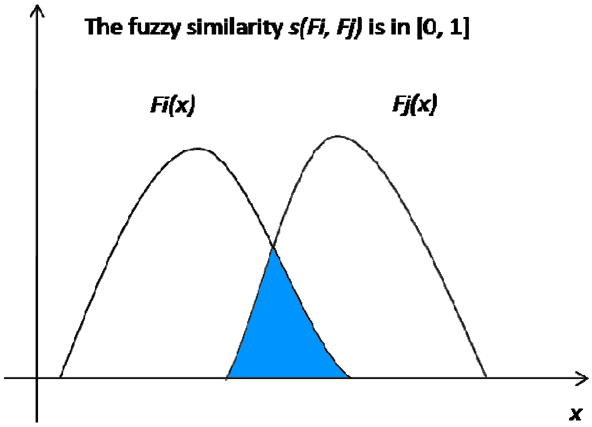


Figure 7.1 The similarity of two fuzzy sets using the fuzzy similarity measure

If the similarity of any two fuzzy subsets for input variables can be suitably controlled, the completeness of fuzzy partitions can be achieved and also the distinguishability can be improved. However, the whole fuzzy system could be incomplete even if the fuzzy partitions of the input are complete. That is to say, the completeness of fuzzy rules also has to be investigated to guarantee the completeness of the fuzzy systems. This issue will be discussed with the compactness of fuzzy systems.

There is another adjustment for fuzzy systems and it is for fuzzy rules, especially for rules generated from data. Here we have to deal with conflicting rules that have the same antecedents but difference consequents. This is called inconsistency of fuzzy rules. If the rules are generated from data mixed with noise, the problem is more serious. In most of the suggestions so far, a degree of belief or strength of rule firing is assigned to each rule and the one with the maximum degree will be accepted to resolve the inconsistency of the generated fuzzy rules.

It is desirable to provide the definition of **consistency** first. As Jin mentioned in [76], [77], fuzzy rules are considered to be inconsistent if

1) fuzzy rules have very similar antecedent variables but rather different consequent variables and
2) they are in conflict with the expert knowledge.

It is possible that two fuzzy rules may be inconsistent when their antecedents are very similar, not necessarily the same. For example, there are two cases of possible inconsistency between two fuzzy rules as shown in Figure 7.2 and 7.3. These examples are extracted from the generated rules on the Fishers' Iris data set using the rough set approach.

R[4]: IF SL=3 AND PL=3 AND PW=3 THEN DECISION=1 [ND]
R[12]: IF SL=3 AND PL=3 AND PW=3 THEN DECISION=2 [ND]

Figure 7.2 The two fuzzy rules with the same antecedents but different consequents

R[2]: IF SL=2 AND PL=3 AND PW=3 THEN DECISION=1 [D]
R[19]: IF SL=0 AND PL=3 AND PW=3 THEN DECISION=2 [D]

Figure 7.3 The two fuzzy rules with similar antecedents but different consequents

It is essential to provide a certain measure of similarity between two rules, since the concept of the consistency is quite abstract. We adapted the definition of the consistency of fuzzy rules using a fuzzy similarity from Jin's proposal in [76], [77], because it is a very good model to consider the similarity of fuzzy rules for the closeness between two rules even when the antecedents are not the same.

The definition of the consistency in [76] is generally suitable for the Mamdani type fuzzy model. But, the proposed system is the T-S type fuzzy system in which the consequent part of fuzzy rules is a function of a linear combination of input variables. If the T-S type consequent part is simplified as a constant, it is possible to use the mathematical equation in [76] to calculate the consistency because a normal fuzzy set can be reduced to a form of a fuzzy singleton. If the consequent is retained as a function, it is difficult to calculate the consistency between rules because of the difficulty to understand the physical meaning of the real functions. This point should be examined further with more theoretical supports.

In general, the generated fuzzy rules from data are quite redundant which means it is not optimal. This leads to a need for the **compactness** for fuzzy systems. In the case of full combination of all input variables to build a fuzzy system, the total number of fuzzy rules is an exponential number which is a computational burden for the system. As mentioned earlier, the compactness of a fuzzy system is strongly required to reduce the system complexity when the number of input features is increased, especially for the T-S type fuzzy model.

As suggested in chapter 5, the rough set approach is applied to the proposed framework of ARFIS to make the generated fuzzy system compact by deploying the knowledge-reduction process without losing its original classification power. Also the theoretical investigation on the generated rules to ensure the full coverage of the input and output relationship of the given information is carried out as proposed earlier. This is to guarantee the completeness of the rule structure and to optimize the T-S type fuzzy rules towards the compactness of the proposed system.

The interpretation of the **flexibility** of a certain system is dependent on the system design or objective. In general, the flexibility for a system is used as a term for a capability of a system to achieve its aim under the different conditions of operations in the environment. Here, the flexibility of fuzzy systems is defined as applicability of a system onto different problems or application domains.

In order to show the viability of the proposed system, we applied our system to different application domains to resolve the different problems. Firstly, the pattern classification scheme is chosen to prove that the proposed system is an excellent tool as a framework of ARFIS by achieving better classification accuracy even on complex higher dimensional data sets. Next, the face recognition task is selected to reduce the huge number of features generated which are from 2D face images and to recognize each person as an identified object using the compact fuzzy rule system. As a result, by deploying the PCA-Rough-Fuzzy system, the number of input features is reduced effectively and the recognition rate is very competitive. For the final application, the mobile robot navigation is chosen to demonstrate better robot navigation behavior in a number of different environments. It is shown that the navigation using the proposed rough-fuzzy system is generally better than other control schemes.

For better **adaptability** of fuzzy systems, a number of approaches have been suggested so far as mentioned earlier. This topic has been one of the popular issues of research in the fuzzy community. A number of different learning approaches have been applied to fuzzy systems, for instance, supervised and unsupervised learning [78], [79], reinforcement learning [80], neural networks-based learning [12], and so forth. In this thesis, the definition of the adaptability is the capability of learning within the context of fuzzy systems.

Regarding the proposed system, we selected the least square estimate and the conjugate gradient descent method for the adaptive mechanism of the ARFIS. To find the coefficients of the consequents of the T-S type fuzzy rules during the system construction with the training sample data, the least square method is utilized. Once the system is established with the training data set, the Polak-Ribiere conjugate gradient function is employed to adjust the parameters of antecedent membership functions towards a smaller RMSE error value.

In a data-driven fuzzy model, one of the biggest issues is the **interpretability** of a fuzzy system. According to the original concept of fuzzy sets theory in Zadeh's paper [1], it is well known that one of the motivations to use fuzzy systems in system modeling is that with a fuzzy system designed using human linguistics it is easy to understand the characteristics of the system behavior. However, the initial system design with good interpretability could be lost after the adaptation process of fuzzy systems. In order to resolve this problem, many approaches have been suggested. For example, the interpretability is controlled by limiting the position of membership functions in [81]. As an alternative, the overlapped and similar membership functions are merged to adjust the fuzzy system to be more interpretable in [82]. For the T-S type fuzzy rules, the interpretability of their consequents is considered in [83] during the local learning process.

The interpretability of fuzzy systems heavily depends on the distribution of the membership functions. The generated fuzzy partition should be complete and distinguishable towards better fuzzy rule generation and more precise meaning of fuzzy subsets. The distinguishability of the fuzzy subsets is the first priority to improve the interpretability of fuzzy systems. There are no clear discussions or definitions so far for the interpretability of a fuzzy system. Also there are no well-established criteria for the distinguishability of fuzzy subsets. The fuzzy similarity-based approach has been discussed in [77] with the regularized learning method to improve the interpretability of a fuzzy system.

For the proposed system, the improvement of the interpretability is being developed to extend this work towards better distinguishability of fuzzy partition of the input domain. There will be more theoretical development for better definition of the interpretability.

List of Publications by Author

C. S. Lee, T. Bräunl, A. Zaknich, "A Rough-Fuzzy Controller for Autonomous Mobile Robot Navigation," The 3rd IEEE International Conference on Intelligent Systems 2006 (IS 2006), pp. 679-682, London, England, Sept. 2006.

C. S. Lee, A. Zaknich, T. Bräunl, "An Adaptive T-S type Rough-Fuzzy Inference System (ARFIS) for Pattern Classification," IEEE International Conference on North American Fuzzy Information Processing Society 2007 (NAFIPS'07), pp. 117-122, San Diego, US, June 2007.

C. S. Lee, A. Zaknich, T. Bräunl, "A Framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS)," IEEE International Conference on Fuzzy Systems 2008 (FUZZ-IEEE2008) in IEEE World Congress on Computational Intelligence 2008 (WCCI 2008), pp. 567-574, Hong Kong, June 2008.

Submitted in-review; C. S. Lee, A. Zaknich, T. Bräunl, "An Adaptive T-S type Rough-Fuzzy Inference System (ARFIS) for Pattern Classification with efficient knowledge-reduction approaches on higher dimensional data," International Journal of Approximate Reasoning (IJAR), Elsevier.

Submitted in-review; C. S. Lee, T. Bräunl, A. Zaknich, "A new Rough-Fuzzy Controller with system performance comparison for Autonomous Mobile Robot Navigation," Journal of Intelligent Systems, Springer.

Bibliography

[1] L. A. Zadeh, "Fuzzy Sets," Information and Control, pp. 338-353, 1965.

[2] S. Horikawa, T. Furuhashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm," IEEE Transactions on Neural Networks, vol. 3, pp. 801-806, Sept. 1992.

[3] J.-Q. Chen, Y.-G. Xi, "Nonlinear System Modeling by Competitive Learning and Adaptive Fuzzy Inference System," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 28, no. 2, pp. 231-238, May 1998.

[4] G. K. I. Mann, R. G. Gosine, "Adaptive hierarchical tuning of fuzzy controllers," Expert Systems, vol. 19, no. 1, pp. 34-45, Feb. 2002.

[5] H. R. Berenji, P. Khedkar, "Learning and tuning fuzzy logic controllers through reinforcements," IEEE Transactions on Neural Networks, vol. 3, no. 5, pp. 724-740, Sept. 1992.

[6] H. B. Kazemian, "Study of learning of fuzzy controllers," Expert Systems, vol. 18, no. 4, pp. 186-193, Sept. 2001.

[7] T. Takagi, M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, pp.116-132, Jan. 1985.

[8] K. C. Ng, M. M. Trivedi, "A Neuro-Fuzzy Controller for Mobile Robot Navigation and Multi-robot Convoying," IEEE Transactions on Systems, Man, and Cybernetics , Part B: Cybernetics, vol. 28, no. 6, pp. 829-840, Dec. 1998.

[9] J.-S. Wang, C. S. George. Lee, "Self-Adaptive Recurrent Neuro-Fuzzy Control of an Autonomous Underwater Vehicle," IEEE Transactions on Robotics and Automation, vol. 19, no.2, pp. 283-295, Apr. 2003.

[10] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," Fuzzy Sets & Systems, vol. 28, pp. 15-33, 1988.

[11] M. Sugeno and K. Tanaka, "Successive identification of a fuzzy model and its application to prediction of complex systems," Fuzzy Sets & Systems, vol. 42, pp. 315-334, 1991.

[12] J.–S. R. Jang, "ANFIS: Adaptive network based fuzzy inference systems," IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665-685, Mar. 1993.

[13] J. J. Buckley and Y. Hayashi, "Fuzzy neural network: A survey," Fuzzy Sets & Systems, vol. 66, pp. 1-13, 1994.

[14] G. M. Dimirovski, I. I. Loevenec, D. J. Tanevska, "Applied Adaptive Fuzzy-Neural Inference Models: Complexity and Integrity Problems," The 2nd IEEE International Conference on Intelligent Systems, pp. 45-52, June 2004.

[15] S. Horikawa, T. Furuhashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm," IEEE Transactions on Neural Networks, vol. 3, pp. 801-806, Sept. 1992.

[16] M.R. Akbarzadeh-T, E. Tunstel, K. Kumbla, M. Jamshidi, "Soft computing paradigms for hybrid fuzzy controllers: experiments and applications," The 1998 IEEE International Conference on Fuzzy Systems in IEEE World Congress on Computational Intelligence (WCCI 98), vol. 2, pp. 1200-1205, 4-9 May 1998.

[17] M. Setnes, H. Roubos, "GA-fuzzy modeling and classification: complexity and performance," IEEE Transactions on Fuzzy Systems, vol. 8, no. 5, pp. 509-522, Oct. 2000.

[18] K. Valarmathi, J. Kanmani, D. Devaraj, T. K. Radhakrishnan, "Hybrid GA Fuzzy Controller for pH Process," International Conference on Computational Intelligence and Multimedia Applications 2007, vol. 4, pp.13-18, 13-15 Dec. 2007.

[19] A. Lotfi, A. C. Tsoi, "Learning Fuzzy Inference Systems Using an Adaptive Membership Function Scheme," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 26, no. 2, pp. 326-331, Apr. 1996.

[20] S. Mitaim, B. Kosko, "The Shape of Fuzzy Sets in Adaptive Function Approximation," IEEE Transactions on Fuzzy Systems, vol. 9, no. 4, Aug. 2001.

[21] Cited In H.-J. Zimmermann, "Fuzzy Set Theory - And its Applications," 3rd Ed, Kluwer Academic Publishers, 1997.

[22] James C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, pp.65-86, 1981.

[23] J. Yu, M.-S. Yang, "A Generalized Fuzzy Clustering Regularization Model With Optimality Tests and Model Complexity Analysis," IEEE Transactions on Fuzzy Systems, vol. 15, no. 5, pp. 904-915, Oct. 2007.

[24] P. M. Kanade, L. O. Hall, "Fuzzy Ants and Clustering," IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 37, no. 5, pp. 758-769, Sept. 2007.

[25] P. Kuang, P. Xiang, L. Chen, "A Novel FCM's Initial Parameters Acquisition Method," The 1st International Multi-Symposiums on Computer and

Computational Sciences 2006 (IMSCCS '06), vol. 2, pp. 224-229, 20-24 April 2006.

[26] L. –X. Wang, “Fuzzy systems are universal approximators,” IEEE International Conference on Fuzzy Systems, pp. 1163-1170, Mar. 1992.

[27] J. J. Buckley, “Sugeno type controllers are universal controllers,” Fuzzy Sets & Systems, vol. 53, pp. 299-303, 1993.

[28] J. L. Castro, and M. Delgado, “Fuzzy systems with defuzzification are universal approximators,” IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 26, no. 1, pp. 149-152, Feb. 1996.

[29] H. Ying, “General Takgi-Sugeno Fuzzy systems are Universal Approximators,” IEEE International Conference on Fuzzy Systems in IEEE World Congress on Computational Intelligence (WCCI 1998), vol.1, pp.819-823, 1998.

[30] H. Ying, “General SISO Takagi-Sugeno Fuzzy Systems with Linear Rule Consequent Are Universal Approximators,” IEEE Transactions on Fuzzy Systems, vol. 6, no. 4, pp. 582-587, Nov. 1998.

[31] H. Ying, Y. Ding, S. Li, and S. Shao, “Comparison of Necessary Conditions for Typical Takagi-Sugeno and Mamdani Fuzzy Systems are Universal Approximators,” IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol. 29, no. 5, pp. 508-514, Sept 1999.

[32] A. Browder, *Mathematical Analysis – An introduction*, Springer-Verlag, pp. 155-158, 1996.

[33] H. Ying, “The Takagi-Sugeno fuzzy controller using the simplified linear control rules are nonlinear variable gain controllers,” Automatica, vol. 34, pp. 157-167, 1998.

[34] H. Ying, “Constructing nonlinear variable gain controllers via the Takagi-Sugeno fuzzy control,” IEEE Transactions on Fuzzy Systems, vol. 6, pp. 226-234, 1998

[35] D. P. Filev and R. R. Yager, “A generalized defuzzification method via BAD distributions,” International Journal of Intelligent Systems, vol. 6, pp. 687-697, 1991.

[36] Z. Pawlak, “Rough Sets,” International Journal of Computer and Information Science, vol. 11, pp. 341-356, 1982.

[37] J. Komorowski, L. Polkowski, and A. Skowron. "Rough sets: A Tutorial". In S.K. Pal and A. Skowron, editors, Rough-Fuzzy Hybridization: A New Method for Decision Making, Springer-Verlag, Singapore, 1998. http://citeseer.ist.psu.edu/komorowski98rough.html

[38] K. Cios, W. Pedrycz, R. Swiniarski, *DATA MINING - Methods for KNOWLEDGE DISCOVERY*, Kluwer Academic Publishers, 1998.

[39] T. Y. Lin and N. Cercone, *Rough Sets and Data Mining – Analysis of Imprecise Data*, Kluwer Academic Publishers, 1997.

[40] S. Tsumoto, H. Tanaka, "Induction of medical expert system rules based on rough sets re-sampling methods," Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care, Journal of the AMIA 1 (supplement), pp. 1066-1070, 1994.

[41] S. Tsumoto, H. Tanaka, PRIMEROSE, "Probabilistic rule induction method based on rough set re-sampling methods," In: Computational Intelligence: An International Journal 11/2, pp. 389-405, 1995.

[42] S. Tsumoto, W. Ziarko, N. Shan, H. Tanaka, "Knowledge discovery in clinical databases based on variable precision rough sets model," In: Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care, New Orleans, Journal of American Medical Informatics Association Supplement, pp. 270-274, 1995.

[43] Z. Pawlak, K. S. lowinski, R. S lowinski, "Rough classification of patients after highly selected vagotomy for duodenal ulcer," Journal of Man-Machine Studies 24, pp. 413-433, 1986.

[44] J. Fibak, Z. Pawlak, K. S lownski, R. S lownski, "Rough sets based decision algorithm for treatment of duodenal ulcer by HSV," Bulletin of Polish Academic Science - Biological Science 34/10-12, pp. 227-246, 1986.

[45] J. Fibak, K. S lownski, R. S lowinski, "The application of rough sets theory to the verification of treatment of duodenal ulcer by HSV," In: Proceedings of the Sixth International Workshop on Expert Systems their Applications, Agence de l'Informatique, Paris, pp. 587-599, 1986.

[46] J.W. Grzyma - Busse, "Applications of the rule induction system LERS," In: Polkowski and Skowron [43], pp. 366-375, 1998.

[47] L. Polkowski, A. Skowron (Eds.), *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, Physica-Verlag, Heidelberg, 1998.

[48] A. Czy_zewski, "Speaker-independent recognition of digits - Experiments with neural networks, fuzzy logic rough sets," Journal of the Intelligent Automation and Soft Computing 2/2, pp. 133-146, 1996.

[49] A. Czajewski, "Rough sets in optical character recognition," In: Polkowski and Skowron [51], pp. 601-604, 1998.

[50] L. Polkowski, A. Skowron (Eds.), The 1st International Conference on Rough Sets and Soft Computing (RSCTC'98), Warszawa, Poland, June 22-27, Springer-Verlag, LNAI 1424, 1998.

[51] A. An, N. Shan, C. Chan, N. Cercone, W. Ziarko, "Discovering rules from data for water demand prediction," Proceedings of the Workshop on Machine Learning in Engineering (IJCAI'95), Montreal, pp. 187-202, 1995; see also, Journal of Intelligent Real-Time Automation, Engineering Applications of Artificial Intelligence 9/6, pp. 645-654, 1995.

[52] J. Catlett, "On changing continuous attributes into ordered discrete attributes," In: Y. Kodrato, (Ed.), Machine Learning-EWSL-91, In: Proc. of the European Working Session on Learning, Porto, Portugal, March 1991, LNAI, pp. 164-178, 1991.

[53] J. W. G.-Busse, "LERS – a system for learning from examples based on rough sets," In: *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory*, R. Slowinski, pp. 3-18, Kluwer Academic Publishers, Boston, 1992.

[54] S. Chanas, D. Kuchta, "Further remarks on the relation between rough and fuzzy sets," Fuzzy sets & Systems, vol. 47, pp.391-394, 1992.

[55] Z. Pawlak, "Rough sets and fuzzy sets," Fuzzy Sets & Systems, vol. 17, pp. 99-102, 1985.

[56] M. Wygralak, "Rough sets and fuzzy sets – some remarks on interrelations," Fuzzy Sets & Systems, vol. 29, pp. 241-243, 1989.

[57] R. Biswas, "On rough sets and fuzzy rough sets," Bulletin of the Polish Academy of Sciences, Mathematics, vol. 42, pp. 345-349, 1994.

[58] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," International Journal of General systems, vol. 17, pp.191-209, 1990.

[59] D. Dubois and H. Prade, "Putting rough sets and fuzzy sets together," In: *Intelligent decision Support: Handbook of Applications and Advances of the rough Sets Theory*, R. Slowinski, Kluwer Academic Publishers, Boston, pp. 203-222, 1992.

[60] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice Hall, 1995.

[61] D. Waillaeys and N. Malvache, "The use of fuzzy sets for the treatment of fuzzy information by computer," Fuzzy Sets & Systems, vol. 5, pp.323-328, 1981.

[62] A. Nakamura, "Fuzzy rough sets," Notes on Multiple-valued Logic in Japan, vol. 9, pp.1-8, 1988.

[63] S. Nanda and S. Maumdar, "Fuzzy rough sets," Fuzzy Sets & Systems, vol. 45, pp. 157-160, 1992.

[64] T. B. Iwinski, "Algebraic approach to rough sets," Bulletin of the Polish Academy of Sciences, Mathematics, vol. 35, pp. 673-683, 1987.

[65] L. I. Kuncheva, "Fuzzy rough sets: application to feature selection," Fuzzy Sets and Systems, vol. 51, pp.147-153, 1992.

[66] Y. Y. Yao, "On combining rough and fuzzy sets," Proceedings of the CSC'95 Workshop on Rough Sets and Database Mining, T. Y. Lin (Ed.), San Jose State University, vol. 9, 9 pages, 1995.

[67] J. C. Bezdek and J. D. Harris, "Fuzzy partitions and relations: and axiomatic basis for clustering," Fuzzy Sets & Systems, vol.1, pp. 111-127, 1978.

[68] A. Nakamura and J. M. Gao, "A logic for fuzzy data analysis," Fuzzy Sets & Systems, vol. 39, pp.127-132, 1991.

[69] P.K. Simpson, "Fuzzy Min-Max Neural Networks–Part 1: Classification," IEEE Transaction on Neural Networks, vol. 3, no.5, pp.776-786, Sept. 1992.

[70] S. Abe and M.S. Lan, "Fuzzy Rules Extraction Directly from Numerical Data for Function Approximation," IEEE Transaction on System, Man, and Cybernetics, vol. 25, no.1, pp.119-129, Jan. 1995.

[71] G.O.A. Zapata, R.K.H. Galvao, and T. Yoneyama, "Extracting Fuzzy Control Rules from Experimental Human Operator Data," IEEE Transaction on System, Man, and Cybernetics, Part B: Cybernetics, vol. 29, no. 3, pp. 25-40, Feb. 1999.

[72] S. K. Pal, S. Mitra, P. Mitra, "Rough-Fuzzy MLP: Modular Evolution, Rule Generation, and Evaluation," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no.1, pp. 14-25, Jan./Feb. 2003

[73] X. Zeng and M. G. Singh, "Approximation Theory of Fuzzy Systems-MIMO Case," IEEE Transactions on Fuzzy Systems, vol. 3, no. 2, pp. 219-235, May 1995.

[74] R. Jensen and Q. Shen, "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 12, pp. 1457-1471, Dec 2004.

[75] G. Jäger and U. Benz, "Measures of Classification Accuracy Based on Fuzzy Similarity," IEEE Transactions on Geo-science and Remote sensing, vol. 38, no. 3, pp. 1462-1467, May 2000.

[76] Y. Jin, W. Seelen, and B. Sendhoff, "On Generating FC^3 Fuzzy Rule Systems from Data Using Evolution Strategies," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 29, no. 6, pp. 829-845, Dec. 1999.

[77] Y. Jin, "Fuzzy Modeling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement," IEEE Transactions on Fuzzy Systems, vol. 8, no. 2, pp. 212-221, Apr. 2000.

[78] C. T. Lin and C. S. G. Lee, "Neural network-based fuzzy control systems," IEEE Transactions on Computer, vol. 40, pp. 1320-1336, Dec. 1991.

[79] J. Nie and D. A. Linkens, "Fast self-learning multivariable fuzzy controllers constructed from a modified CPN network," International Journal of Control, vol. 60, no. 3, pp. 369-393, 1994.

[80] H. R. Berenji and P. Khedkar, "Learning and tuning fuzzy controllers through reinforcement," IEEE Transactions on Neural Networks, vol. 3, pp. 724-739. Sept. 1992.

[81] A. Lotfi, H. C. Anderson, and A. C. Tsoi, "Interpretation preservation of adaptive fuzzy inference systems," International Journal of Approximation Reasoning, vol. 15, no.4, 1996.

[82] M. Setnes, R. Babuska, U. Kaymak, and H.R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 28, pp. 376-386, June 1998.

[83] J. Yen, L. Wang, and W. Gillespie, "A global-local learning algorithm for identifying Takagi-Sugeno-Kang fuzzy models," In Proc. IEEEE International Conference on Fuzzy Systems, , pp. 967-972, Anchorage, AK, May 1998.

[84] S. Fahlman and C. Lebiere, "The Cascade-Correlation Learning Architecture," Carnegie Mellon Univ., School of Computer Science, Technical Report CMU-CS-90-100, Feb. 1990.

[85] T-P. Hong and S.-S. Tseng, "A Generalized Version Space Learning Algorithm for Noisy and Uncertain Data," IEEE Transaction on Knowledge and Data Engineering, vol. 9, no. 2, pp. 336-340, Mar.-Apr. 1997.

[86] S.C. Newton, S. Pemmaraju, and S. Mitra, "Adaptive Fuzzy Leader Clustering of Complex Data Sets in Pattern Recognition," IEEE Transaction on Neural Networks, vol. 3, no.5, pp.794-800, Sept. 1992.

[87] T.P. Wu and S.M. Chen, "A New Method for Constructing Membership Functions and Fuzzy Rules from Training Examples," IEEE Transaction on System, Man, and Cybernetics – Part B: Cybernetics, vol. 29, no.1, pp.25-40, Feb. 1999.

[88] R. Setiono, “Extracting M-of-N Rules from Trained Neural Networks,” IEEE Transaction on Neural Networks, vol. 11, no. 2, pp.512-519, Mar. 2000.

[89] B.C. Lovel and A.P. Bradley, “The Multi-scale Classifier,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 18, no. 2, pp. 124-137, Feb. 1996.

[90] H.-M. Lee, C.-M. Chen, J.-M. Chen, and Y.-L. Jou, “An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy,” IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 31, no. 3, pp.426-432, June 2001.

[91] A. Chatterjee and A. Rakshit, “Influential Rule Search Scheme (IRSS) – A New Fuzzy Pattern Classifier,” IEEE Transaction On Knowledge and Data Engineering, vol. 16, no. 8, pp. 881-893, Aug. 2004.

[92] C.L. Blake and C.J. Merz, “UCI Repository of Machine Learning Databases,” University of California, Irvine, Department of Information and Computer Science, http://www.ics.uci.edu/~mlearn /MLRepository.html, 1998.

[93] C.S. Lee A. Zaknich, T. Bräunl, “An Adaptive T-S type Rough-Fuzzy Inference System (ARFIS) for Pattern Classification,” IEEE International Conference on North American Fuzzy Information Processing Society 2007 (NAFIPS’07), pp. 117-122, June 2007.

[94] Z-X. Yin, J-H. Chiang, “Patterns Discovery on Complex Diagnosis and Biological Data Using Fuzzy Latent Variables,” The 23rd IEEE International Conference on Data Engineering (ICDE 2007), pp. 576-585, April 2007.

[95] K. Saastamoinen, J. Ketola, “Medical Data Classification using Logical Similarity Based Measures,” IEEE International Conference on Cybernetics and Intelligent Systems, pp. 1-5, June 2006.

[96] M. Turk and A. Pentland, “Face Recognition Using Eigenfaces,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, June, 1991.

[97] M. Turk and A. Pentland, “Eigenfaces for Recognition,” Journal of Cognitive Neuro-science, March. 1991.

[98] K-C. Kwak and W. Pedrycz, “Face Recognition Using an Enhanced Independent Component Analysis Approach,” IEEE Transactions on Neural Networks, vol. 18, no. 2, pp. 530-541, Mar. 2007.

[99] L. H. Bac and N. A. Tuan, “Using Rough Set in Feature Selection and Reduction in Face Recognition Problem,” Lecture Notes in Computer Science, SpringerLink, vol 3518/2005, pp. 226-233, 2005.

[100] MIT Media Labs - ftp://whitechapel.media.mit.edu/pub/images/

[101] C. C. Lee, “Fuzzy Logic in Control Systems: Fuzzy Logic Controller Part I & II,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 20, no. 2, pp. 404-418 (I), 419-435 (II), Mar./Apr., 1990.

[102] J.-Q. Chen, Y.-G. Xi, “Nonlinear System Modeling by Competitive Learning and Adaptive Fuzzy Inference System,” IEEE Transactions On Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 28, no. 2. pp. 231-238, May 1998.

[103] M. Sarkar and B. Yegnanarayana, “Rough-Fuzzy Membership Functions,” IEEE International Conference on Fuzzy Systems in WCCI 1998, pp. 796- 801, May 1998.

[104] M. Sarkar, “Rough-fuzzy functions in classification,” Journal of Fuzzy Sets and Systems, vol. 132, pp. 353~360, 2002.

[105] Andreas Koestler, Thomas Bräunl, “Mobile Robot Simulation with Realistic Error Models,” International Conference on Autonomous Robots and Agents, ICARA 2004, Palmerston North, New Zealand, pp. 46-51, Dec. 2004.

[106] Thomas Bräunl, *Embedded Robotics: Mobile Robot Design and Applications with Embedded Systems*, Springer-Verlag, Berlin, Heidelberg, 2nd Ed. 2006.

[107] M. Mucients and J. Cashillas, “Obtaining a Fuzzy Controller with High Interpretability in Mobile Robots Navigation,” IEEE International Conference on Fuzzy Systems, pp.1637-1642, July, 2004.

[108] Franklin, G.F. and J.D. Powell, *Digital control of Dynamic Systems*, Addison Wesley, 1990.

[109] Dahlin, E.B, “Designing and tuning digital controllers,” Instrumentation Technology, 41, pp.77~83. 1970.

[110] C. S. Lee, T. Bräunl, A. Zaknich, “A Rough-Fuzzy Controller for Autonomous Mobile Robot Navigation,” The 3rd IEEE International Conference on Intelligent Systems 2006 (IS 2006), pp. 679-682, Sept. 2006.

[111] *Submitted in-review;* C. S. Lee, T. Bräunl, A. Zaknich, “A new Rough-Fuzzy Controller with system performance comparison for Autonomous Mobile Robot Navigation,” Journal of Intelligent Information Systems, Springer.

[112] I. Kamon and E. Rivlin, "Sensory-Based Motion Planning with Global Proofs," IEEE Transactions on Robotics and Automation, vol. 13, no. 6, pp. 814-822, Dec, 1997.

[113] C. S. Lee, A. Zaknich, T. Bräunl, "A Framework of Adaptive T-S type Rough-Fuzzy Inference Systems (ARFIS)," IEEE International Conference on Fuzzy Systems 2008 (FUZZ-IEEE2008) in IEEE World Congress on Computational Intelligence 2008 (WCCI 2008), pp. 567-574, Hong Kong, June 2008.

[114] *Submitted in-review;* C. S. Lee, A. Zaknich, T. Bräunl, "An Adaptive T-S type Rough-Fuzzy Inference System (ARFIS) for Pattern Classification with efficient knowledge-reduction approaches on higher dimensional data," International Journal of Approximate Reasoning (IJAR), Elsevier.

[115] Jun Shao and Dongsheng Tu, *The jackknife and bootstrap*, Springer-Verlag, New York, US, 1995.