

# CLASSIFICATION OF HAND-WRITING IDENTIFICATION BASED ON ANFIS

LEONG SHU CHING



UNIVERSITI TEKNIKAL MALAYSIA MELAKA



**CLASSIFICATION OF HAND-WRITING IDENTIFICATION BASED ON ANFIS**

**LEONG SHU CHING**

**This report is submitted in partial fulfillment of the requirements for the Bachelor of  
Computer Science (Artificial Intelligence)**

**FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY  
UNIVERSITI TEKNIKAL MALAYSIA MELAKA  
2011**



## **DEDICATION**

Specially

To my beloved parent and family members,

And not forgetting to all my friends and supervisors

who have provided a lot of love, guidance and encouragement

during the completion of report.

## ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Dr. Azah Kamilah Binti Draman @ Muda for her guidance, encouragement, suggestions and continuous support throughout my project. Without her valuable guidances, this project would not be possible.

Thanks also to Dr. Choo Yun huoy for giving some idea about the project, explain to me about fuzzy logic part and giving me a lot of encouragements to complete the project.

I am grateful to Universiti Teknikal Malaysia Melaka (UTeM) for offering the Project Sarjana Muda (PSM) as a compulsory subject for students. By the way complete the project, I have learned a lot of knowledge and understand my weakness.

Not forgetting to all my family and friends for their supports and encouragement during accomplish my project. Without their continued support, this project would not have been realized.

Last but not least, my gratitude to all who were helping me along the way. Thank you all.

## ABSTRACT

Hand-writing identification is an active topic in the image processing and pattern recognition, though a lot affords have put in, it still remain a challenge issue. This project was focused on classification of hand-writing identification. Adaptive-Network-Based Fuzzy Inference System (ANFIS) is the choosing classification technique to identify the hand-writing. ANFIS is a Sugeno type Neuro-fuzzy model, where it is the combination of neural network and fuzzy logic. In this project, an experiment was carried out where several testing are done to produce the accurate classification performance. The testing involved the choosing of Neuro-fuzzy system, three classifiers have been selected: ANFIS-GRID, ANFIS-SUB, and ANFIS\_FCM. Besides, data type also tested in the experiment to check whether the discretized data or original data show the better result. Last, the cross-validation technique also being used in experiment to set the training data set and testing data set. The performance of ANFIS was evaluated in terms of training performance and classification accuracy. The discretization data and ANFIS-FCM show the better result compared to the others.

## ABSTRAK

Pengenalan penulis tangan adalah salah satu topik yang hangat dalam bidang imej pemprosesan dan corak pengiktirafan, namun banyak usaha digunakan dalam bidang ini, pengenalan penulis tangan masih adalah satu cabaran yang hebat. Maka, projek ini telah memberi tumpuan kepada klasifikasi pengenalan. Adaptive-Network-Based fuzzy inference system (ANFIS) adalah klasifikasi teknik yang akan digunakan dalam projek ini. ANFIS terdiri daripada Sugeno type neuro fuzzy model, gabungan rangkaian neural dan logik kabur. Dalam projek ini, beberapa ujian telah dijalankan untuk menghasilkan prestasi pengelasan yang tepat. Ujian ini melibatkan penggunaan sistem neuro-kabur, tiga jenis pengelasan telah dipilih iaitu ANFIS-GRID, ANFIS-SUB, dan ANFIS-FCM. Selain itu, jenis data juga dikaji dalam eksperimen. Tujuannya adalah untuk menepatpastikan mana salah satu jenis data sesuai digunakan dalam eksperimen. Akhir sekali, cross-validation juga digunakan dalam eksperimen untuk menyediakan data latihan dan data ujian. Prestasi ANFIS telah dinilai dari segi prestasi latihan dan ketepatan pengelasan. Discretized data dan ANFIS\_FCM menunjukkan hasil yang lebih baik berbanding dengan yang lain.



## TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	II
	DEDICATION	III
	ACKNOWLEDGEMENTS	IV
	ABSTRACT	V
	ABSTRAK	VI
	TABLE OF CONTENTS	VII
	LIST OF TABLES	X
	LIST OF FIGURES	XI
	LIST OF ABBREVIATIONS	XIII
	LIST OF APPENDICES	XIV
I.	INTRODUCTION	
	1.1 Project Background	1
	1.2 Problem Statement	2
	1.3 Objective	2
	1.4 Scope	3
	1.5 Project significance	3
	1.6 Expected Output	4

1.7	Conclusion	4
II.	LITERATURE REVIEW AND PROJECT METHODOLOGY	
2.1	Introduction	5
2.2	Facts and Findings	
2.2.1	Hand-writing domain	5
2.2.2	Existing Works	6
2.3	Research Methodology	7
2.4	Project Requirements	10
2.5	Project Schedule and Milestone	10
2.6	Conclusion	11
III.	ARCHITECTURE AND EXPERIMENT SET UP	
3.1	Introduction	12
3.2	Neuro-Fuzzy model	12
3.3	Adaptive Neuro-Fuzzy Inference (ANFIS)	13
3.3.1	Architecture design of ANFIS	14
3.3.2	Basic Learning Algorithm	16
3.3.3	Techniques	19
3.3.4	Basic Flow Diagram of ANFIS	25
3.4	Conclusion	26
IV.	EXPERIMENTAL RESULT	
4.1	Introduction	27
4.2	Experimental Setup	27
4.2.1	Data Preparation	28
4.2.2	Fis Generation	29
4.2.3	Training Phase	30
4.3	Conclusion	31

V.	RESULTS AND DISCUSSION	
5.1	Introduction	32
5.2	Result Analysis	32
5.2.1	Cross validation	33
5.2.2	Data Preprocessing	35
5.2.3	Classifiers	37
5.2.4	Training and Testing Error	38
5.2.5	FIS Structure	40
5.3	Conclusion	47
VI.	CONCLUSION AND FUTURE WORKS	
6.1	Conclusion	48
6.2	Future Works	49
	REFERENCES	51
	Appendices A-B	53

**LIST OF TABLES**

<b>TABLE</b>	<b>TITLE</b>	<b>PAGE</b>
4.1	The parameter setting for ANFIS-SUB	30
5.1	The accuracy of classification for overall cross-validation	33
5.2	The average performance rating for discretized and original data	35
5.4	Training error and testing error	39

**LIST OF FIGURES**

FIGURE	TITLE	PAGE
2.1	Flow diagram of research methodology	7
2.2	The flow chart of ANFIS model	9
3.1	The ANFIS architecture	14
3.2	Steps for LSE algorithm	18
3.3	Steps for Steepest Descent Learning Algorithm	19
3.4	Typical the fuzzy Membership functions.	21
3.5	The diagram of FIS	21
3.6	Steps of Fuzzy reasoning	22
3.7	The basic flow diagram of ANFIS	25
5.1	Result of cross-validation with original data	34
5.2	Result of cross-validation with discretized data.	34
5.3	Comparison between discretized data and original data	36
5.4	The statistic result of performance rating	37
5.5	The average results of training error and testing error.	39

5.6	The Rule Viewer of ANFIS-GRID	42
5.7	The Rule Viewer of ANFIS-SUB	41
5.8	The Rule Viewer of ANFIS-FCM	42
5.9	The ANFIS Model structure of ANFIS-GRID	43
5.10	The ANFIS Model structure of ANFIS-SUB	44
5.11	The ANFIS Model structure of ANFIS-FCM	44
5.12	The Rule Editor of ANFIS-GRID	45
5.13	The Rule Editor of ANFIS-SUB	46
5.14	The Rule Editor of ANFIS-FCM	46

**LIST OF ABBREVIATIONS**

ANFIS	-	Adaptive Neuro-Fuzzy Inference System
ANFIS-GRID	-	Grid Partition based ANFIS
ANFIS-FCM	-	Fuzzy C-means based ANFIS
ANFIS-SUB	-	Subtractive clustering based ANFIS
ANN	-	Artificial Neural Network
FCM	-	Fuzzy C-Means
FIS	-	Fuzzy Inference System
MF	-	Membership Function
NN	-	Neural Network

## CHAPTER I

### INTRODUCTION

#### 1.1 Project Background

Data mining is the process of extracting patterns from large data sets by analyzing data from different perspectives and summarizing it into useful information. Data mining will apply certain techniques to data with the intention of uncovering hidden pattern and assist in the analysis of collections of observations of behaviour. Data mining is seen as an increasing important tool where it allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified.

Classification is the most common data mining techniques for finding the hidden pattern in data. In classification or clustering, the most important issue is deciding what criteria to classify against. Nevertheless, a criterion for classification must be prepared before separate the data into describable class. As is often the case in classification studies about the number and kind of features and the type of classification criteria are choices.

In this project, a fuzzy-neuro classification method, Adaptive Neuro Fuzzy Inference System, ANFIS, is selected to test the classification result in the hand-writing identification domain. Three neuro-fuzzy systems which are grid partition based fuzzy inference system (FIS), ANFIS-GRID, sub-clustering based FIS, ANFIS-SUB, and Fuzzy C-mean based FIS, ANFIS-FCM, will be presented in this project to test which are more suitable use in hand-writing identification domain.



## 1.2 Problem Statement

Hand-writing identification is an active topic for several decades in the image processing and pattern recognition. Though a lot of efforts have put in this domain, it still remain a challenging issue. Furthermore, hand-writing identification is inapplicable in many important practical applications, for example, the identification of the writers of achieved handwritten documents, crime suspect identification in forensic science. Although hand-writing can be identified by graphologist, it is costly and prone to fatigue. Therefore, a classification technique based on this domain is needed to accurately identify the hand-writer.

Adaptive Neuro-Fuzzy inference system, ANFIS, is combination method which works similarly to that of neural network. It provide a adaptive learning techniques to fuzzy modelling procedure to learn the information about the data set. This neuro-fuzzy system technique can well being used to solve the classification problem. Thus ANFIS is used in this project to classify the writer based on a data set.

## 1.3 Objective

- i. Writer identify :
  - To apply the ANFIS classification method in hand-writing identification data set, in order to identify the writer through information generated in data.
- ii. Experiment :
  - Used three different type of classifiers: ANFIS-GRID, ANFIS-SUB, ANFIS-FCM to generate fuzzy rule.
  - Used the discretization data set and original data set to test the classification performance.

## 1.4 scope

- i. Domain :
  - Hand-writer identification data set will be used in the hold project.
- ii. Classification method:
  - ANFIS is the choosing method for this project.
- iii. Comparison techniques
  - Grid partition, sub-clustering, and fuzzy c-mean are three of the rule generate techniques where will be used to test which are most suitable to be use in writer identification domain.
- iv. Development tool:
  - Matlab will be used as developer tool.
- v. Data preparation:
  - The original data set contain 60 classes, it will take more time to run the experiment, thus data set will be resampled to 10 classes.

## 1.5 Project significance

Handwriting identification as biometric domain has had the interest of researchers for long time, but recently it has been enjoying new interest due to an increased need and effort to deal with probelm ranging from white-collar crime to terrorist threats. In forensic writer identification, pieces of handwriting are compared to identify the writer. These questioned documents can be threatening letters, fraud letters or suspicious suicide notes. Traditionally, this writing identification is performed by human forensic document expert.

## **1.6 Expected Output**

ANFIS can be apply in hand-writer identification domain and FCM will be the best method to generate the fuzzy rule.

## **1.7 Conclusion**

In this chapter, a briefing is given on the overview of the study. This includes project background, problem statements, objective, scope, project significant and the expected output of project. As a conclusion, ANFIS method will be implemented in the hand-writer identification domain and a comparison between classifiers will be carried out.

The next activities that will be developed are Chapter II which is Literature Review and Project Methodology. This chapter includes the Introduction, Facts and Findings, Project Methodology, Project Requirements; Project Schedule and Milestones.

## CHAPTER II

### LITERATURE REVIEW AND PROJECT METHODOLOGY

#### 2.1 Introduction

This chapter will discuss on the literature review and project methodology. Literature review is a discussion of published information in a particular subject area, it can be a simple summary of the sources, but it usually has an organizational pattern and combines both summary and synthesis. Literature review encompasses and analysis on the techniques used and study on the project domain, it will provide a handy guide to the topic and give an overview concept to the research study. It is involves searching, collecting, studying and analyzing the relevant sources such as references books, articels, journals, webpages and others.

Project methodology mean the methodology to manage the projects from start to end. It is needed to steer project in the right direction and keep them on track. In this part will clearly identifies the methodology to be adapted in the project development and it will explain in more details about the activities in every stages of the project. Furthermore, its lists all the tools, software, operating system and hardware used for the project.

## **2.2 Facts and findings**

### **2.2.1 Hand-writing identification**

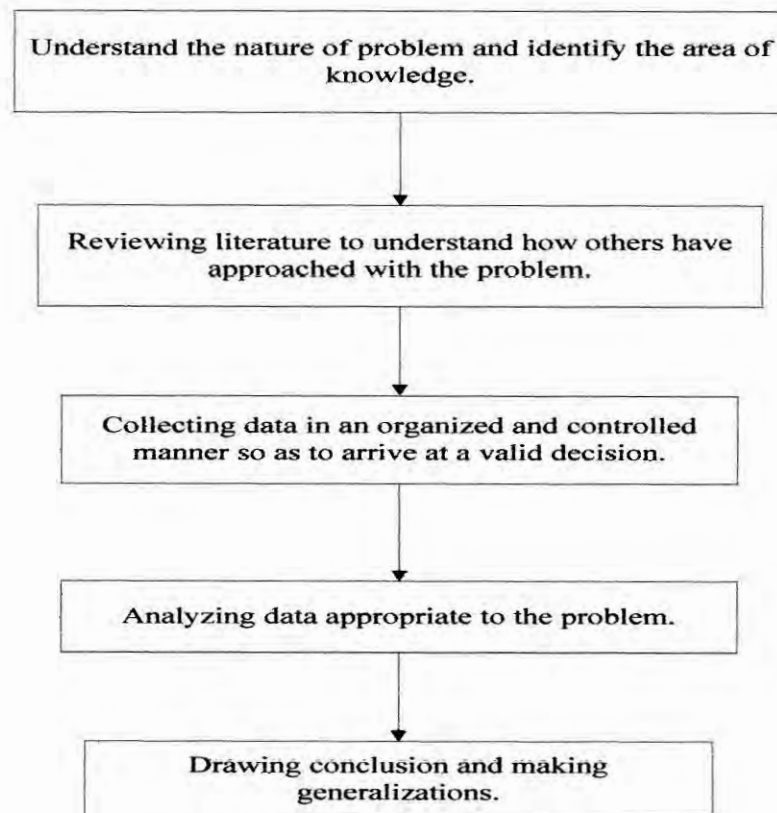
This is a process where a system can identify the hand-writing based on the features provided in a data set. In these data set contain certain features which have important data that system can use to interpret useful information for hand-writing classification. A comparison between classifiers is being carried out. Three of the selected classifiers are Grid Partition, Sub-clustering, and Fuzzy C-mean. Three of these classifiers are tested and compared which are suitable to be used in the ANFIS model for hand-writing identification domain. Besides, two types of data set: discretization and original data set have been used to test the performance of classification.

### **2.2.2 Existing Works**

Adaptive Network based Fuzzy Inference System (ANFIS) is a hybrid system by the combination of neural network and fuzzy system. It was initially introduced by Prof. J. S. Roger Jang in the year of 1993 as state in (Jang, J.-S. R. , 1993). From the research paper by (Azar, et al), (Guo, X.T.,2008) and (Othman.M.F.,2007) have proved that ANFIS model is one of most successful schemes and achieve great success in a wide range of scientific applicants like biometric, control system, signal processing and so on. The accuracy of classification that generated by ANFIS is about 70%. However, ANFIS technique is not yet being used in hand-writing identify domain. Therefore, an experiment is designed to carry out by applying the ANFIS technique in handwriter identification domain for the purpose to test ANFIS work in this domain or not. Furthermore, three ANFIS based neuro-fuzzy system which are ANFIS-GRIB, ANFIS-SUB, and ANFIS-FCM will be use as a testing parameters to compare the resultant performance. An analysis will be done based on the result to determine which method are most suitable to be used in handwriter identification domain.

### 2.3 Research Methodology

To start doing a research, an understanding and overview of research should be built in mind. Research means the search of knowledge, in order to discover and interpret for facts or truth about a subject. There are two ways to do research, one is based on purpose and another is based on method employed in research. Generally, research is systematic because it follows a structural process and step by step in a logical order. The steps involve are show in figure 2.1 flow diagram of research methodology.



**Figure 2.1 Flow diagram of research methodology**

After having an overview with the concept of research, it should keep on with the literature review. It is actually the reading of the works of others before start doing research which can help in identifying the relevance of the research. To carry out an

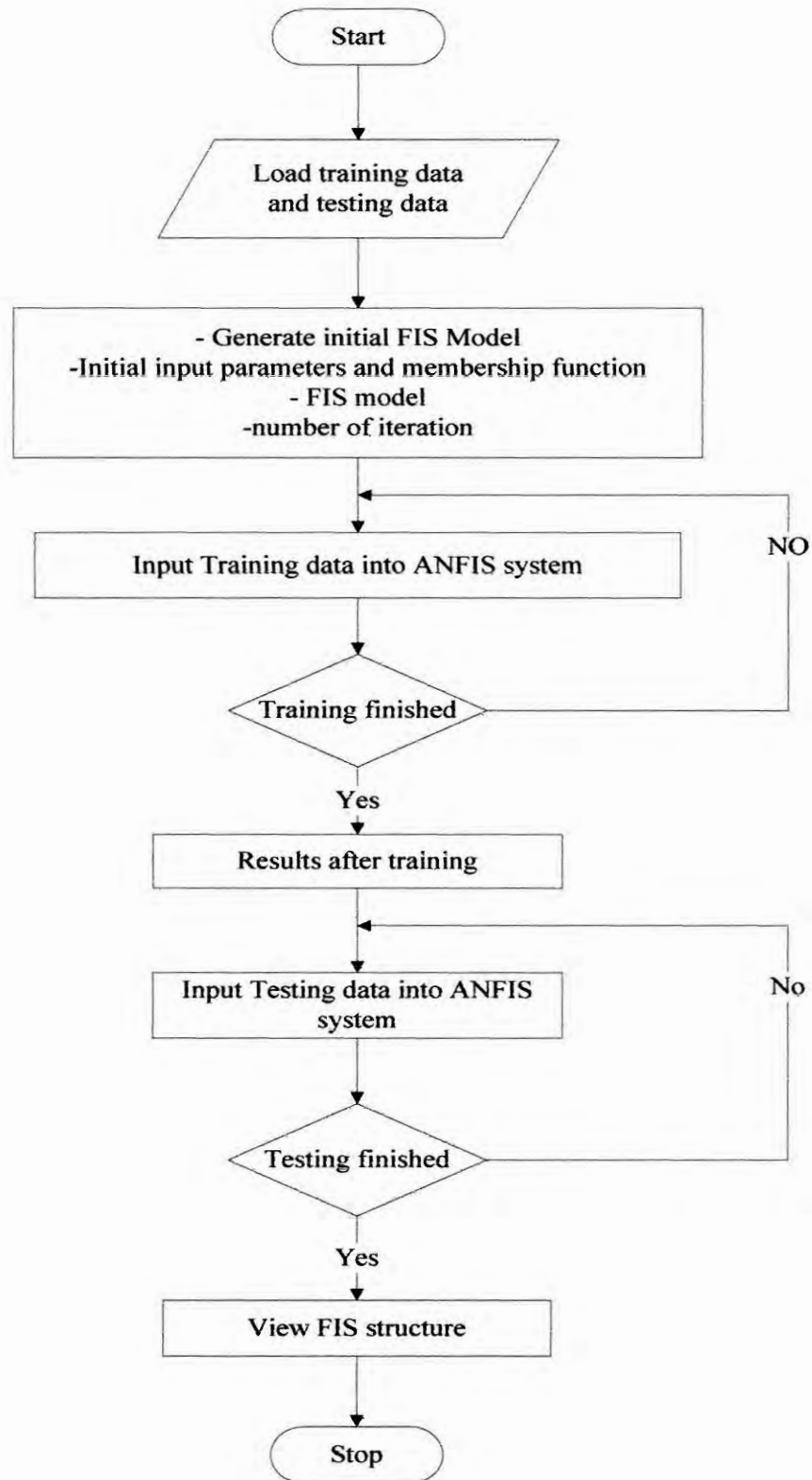
effective literature review (Khalid M.B., Research Methodology. Retrieved on April 2011 from

[http://www.cairo.utm.my/webcairo/download/research\\_method/Research%20Methodology%201.pdf](http://www.cairo.utm.my/webcairo/download/research_method/Research%20Methodology%201.pdf))

1. Should collect a list of source word that relate to research problem.
2. Consultations
3. Using key words check the initial sources for references
4. Check secondary sources (textbok, articles) for relevance information.
5. Check primary sources for information on researh design and methodology.
6. Contact the person who have conducted research in the area.

The next steps for doing research is started selecting and defining a research problem. The criteria for selecting a research problem should base on interest, size of scope, cost, capability, limitation and uniqueness. After select a problem, then should identify the variable and evaluating the problem.

Once the problem situation is located, start conducting with the research which involve five main steps. First is data preparation, two type of data sets will be test in the experiment where one is discretized data and another one is non-discretized. So, to test on the data type, data preprocessing is needed before start the experiment. Besides, the training data set and testing data set are need to prepare as well. The next step is using three different type of neuro-fuzzy model to generate the Fuzzy inference system (FIS), where the neuro-fuzzy model use different method to generate the fuzzy rule. After generate the FIS, the ANFIS model will start use the training data set to learn the algorithm from the fuzzy rule. When training phase is done, the testing data set is required to test the training result. The output will produce after testing phase is done. The overall steps for carried out the experiment has shown in the Figure 2.2. The last step of research is analysis the performance result and make the conclusion for the experiment. Finally, write a report and thesis for the research.



**Figure 2.2: The flow chart of ANFIS model**