



UNIVERSITI PUTRA MALAYSIA

**GRID-BASED CLASSIFIER AS A REPLACEMENT FOR MULTICLASS
CLASSIFIER IN A SUPERVISED NON-PARAMETRIC APPROACH**

MAJID REZA MOHEB POUR

FK 2009 83



**GRID-BASED CLASSIFIER AS A REPLACEMENT FOR MULTICLASS
CLASSIFIER IN A SUPERVISED NON-PARAMETRIC APPROACH**

By

MAJID REZA MOHEB POUR

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in
Fulfilment of the Requirement for the Degree of Master of Science**

October 2009



DEDICATION

I dedicate this dissertation to

My Father,

who taught me how to know myself

and strive to be a better man,

to

The Memory of my Mother,

who taught me all I need to know about

Love, Sacrifice, and Loyalty,

both during her life and after it

(She lives in my heart forever),

and to

My Brothers and Beloved Sisters,

as a small token of my great appreciation.



Abstract of thesis to be presented to the Senate of Universiti Putra Malaysia in
fulfilment of the requirement for the degree of Master of Science

**GRID-BASED CLASSIFIER AS A REPLACEMENT FOR MULTICLASS
CLASSIFIER IN A SUPERVISED NON-PARAMETRIC APPROACH**

By

MAJID REZA MOHEB POUR

October 2009

Chairman: Associate Professor Dr. Adznan Bin Jantan, PhD

Faculty: Engineering

Pattern recognition/classification has received a considerable attention in engineering fields. In most applications, it is desirable to maintain the classification accuracy, but also reduce the classification time. The quality of a given classification technique is measured by the computational complexity, execution time of algorithms, and the number of patterns that can be classified correctly despite any distribution. In this thesis, a new method known as Grid Based Classifier was proposed. This method carries the advantages of the two previous methods in order to improve the classification tasks. The problem with the current lazy algorithms is that they learn quickly, but classify very slowly. On the other hand, the eager algorithms classify quickly, but they learn very slowly. The two algorithms were compared, and the proposed algorithm was found to be able to both learn and classify quickly. The



method was developed based on the grid structure, whereby it was done to create a successful method of improving performance in classification. In the current research, the new algorithm was tested and applied to the multiclass classification of two or more categories, which are important for handling problems related to practical classification. The new method was also compared with the Levenberg-Marquardt back-propagation neural network in the learning stage and the Condensed nearest neighbor in the testing stage to examine the performance of the model. The experimental results on artificial data sets and real-world data sets (from UCI Repository) show that the new method could improve both the efficiency and accuracy of pattern classification. In real-world experiment (Haberman data set), new method allows 1% improvement in training accuracy and 1.8% improvement in testing accuracy and also allows considerable improvement in running time comparing to neural network method.



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Master Sains

**GRID-BASED CLASSIFIER SEBAGAI PENGGANTI PENGKELASAN
BERBAGAI KELAS BAGI PENDEKATAN TIDAK BERPARAMETER
YANG TERKAWAL**

Oleh

MAJID REZA MOHEB POUR

Oktober 2009

Pengerusi: Profesor Madya Dr. Adznan Bin Jantan, PhD

Fakulti: Kejuruteraan

Pengecaman/pengklasifikasian paten telah menjadi tumpuan di dalam bidang kejuruteraan. Di dalam kebanyakan aplikasi pengklasifikasian paten, ketepatan klasifikasi harus dikekalkan dan masa klasifikasi harus dikurangkan. Kualiti kepada teknik klasifikasi diukur melalui kerumitan pengiraan yang digunakan, masa yang diperlukan untuk memproses algoritma, dan jumlah paten yang boleh diklasifikasi dengan tepat tanpa mengira jenis taburan pada data. Tesis ini mencadangkan satu kaedah baru dikenali sebagai Klasifikasi Grid Dasar. Kaedah ini mengadaptasi kelebihan dua kaedah sedia ada untuk memperbaiki tugas-tugas pengklasifikasian. Permasalahan yang timbul dari penggunaan algoritma '*lazy*' adalah pengadaptasian pantas tetapi pengklasifikasian perlahan manakala bagi algoritma '*eager*' pula adalah sebaliknya. Perbandingan dilakukan ke atas kedua-dua algoritma tersebut dan algoritma yang dicadangkan didapati berkebolehan mengadaptasi dan mengklasifikasi dengan



pantas. Kaedah yang dibentuk adalah berdasarkan struktur grid yang bertujuan untuk menghasilkan satu kaedah yang berjaya meningkatkan prestasi pengklasifikasian. Dalam kajian yang dijalankan, algoritma tersebut telah diuji dan digunakan dalam pengklasifikasian kelas berganda terhadap dua atau lebih kategori yang penting dalam menghadapi masalah berkaitan dengan pengklasifikasian praktikal. Kaedah baru tersebut dibandingkan dengan kaedah *Levenberg-Marquardt back-propagation neural network* dalam fasa pembelajaran dan kaedah *Condensed nearest neighbour* dalam fasa generalisasi untuk mengkaji kadar prestasi model tersebut. Hasil kajian terhadap set data buatan dan data sebenar (daripada UCI Repository) menunjukkan method baru ini dapat meningkatkan kadar keberkesanan dan ketepatan pengklasifikasian paten. Dalam uji kaji sebenar (*Haberman* data sebenar), kaedah baru ini membenarkan 1% peningkatan dalam ketepatan latihan dan 1.8% peningkatan dalam ketepatan ujian dan juga membenarkan peningkatan yang besar dalam masa pemprosesan berbanding kaedah *neural network*.

ACKNOWLEDGEMENTS

First of all, I would like to thank the Lord for his enormous support during my life, and then I wish to thank a number of people who in various ways made my graduate studies possible.

I would like to express my gratitude to the supervisory committee members, Associate Professor Dr. Adznan Bin Jantan and Dr. M. Iqbal Bin Saripan for proposing this challenging project and for their confidence in my abilities and also for their great guide and help throughout the work.

I would like to gratefully acknowledge my family specially my father and my brothers Saeedreza & Vahidreza and sisters Fatemeh & Masoomah for their great help, patience, care, guidance, enormous support and understanding of family time lost during my studies. I would not have passed through these years without them.

In addition, special thanks are due to my great friends for their invaluable comments, consultation, guidance and support that kept me on the right track during the work.

Last but not least, I wish to thank head of Computer and Communication department. As well I wish to thank the Deputy Dean of Graduate School in engineering faculty Dr. Norman Mariun and also thank the Deputy Dean and all the staffs of Graduate School for their nice cooperation during my years at University Putra Malaysia.



APPROVAL

I certify that an Examination Committee met on 2008 to conduct the final examination of Majid Reza Mohebpour on his Master degree thesis titled “Grid Based Classifiers for Multiclass Classifier in Supervised Non-parametric Approach” in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Abdul Rahman B. Ramli, PhD

Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Md. Nasir Sulaiman, PhD

Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Mohd. Fadlee A. Rasid, PhD

Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Dzulkifli Bin Mohamad, PhD

Professor
Faculty of Computer Science and Information systems
Universiti Teknologi Malaysia
(External Examiner)

BUJANG KIM HUAT, PhD

Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:



This thesis was submitted to the Senate of University Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Adznan Bin Jantan, PhD

Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

M. Iqbal Bin Saripan, PhD

Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Member)

HASANAH MOHD GHAZALI, PhD

Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 14 January 2010



DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

Majid Reza Mohebpour

Date:



TABLE OF CONTENTS

	Page
DEDICATION	ii
ABSTRACT	iii
ABSTRAK	v
ACKNOWLEDGEMENTS	vii
APPROVAL	viii
DECLARATION	x
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvii
CHAPTER	
1 INTRODUCTION	
1.1 Background	1
1.1.1 Classification	2
1.1.2 Supervised Parametric Methods	3
1.1.3 Supervised Non-parametric Methods	5
1.1.4 Neural Network	6
1.1.5 Noise	7
1.1.6 Decision Boundary	8
1.1.7 Multiclass Classification	11
1.1.8 Grid Structure	12
1.2 Statement of the Problem	13
1.3 Research Objectives	16
1.4 Research Scope	16
1.5 Limitation	17
1.6 Thesis Organization	18
2 LITERATURE REVIEW	
2.1 Nonparametric Classifiers	20
2.2 Parzen Windows	20
2.3 Nearest Neighbor Method	23
2.3.1 Condensed Nearest Neighbor rule	28
2.4 Decision Trees	29
2.5 Neural Network	32
2.5.1 Characteristics	33
2.5.2 Description	34



3	MATERIAL AND METHODS	
3.1	Design of Patter Recognition Systems	44
3.2	Preprocessing and Feature Selection	45
3.3	Classification Model	46
3.4	Grid Based Classifier	47
	3.4.1 Preprocessing	47
	3.4.2 Learning	48
	3.4.3 Testing	55
3.5	Implementation in MATLAB	57
4	RESULTS AND DISCUSSION	
4.1	Experimental Results for Artificial Datasets	64
4.2	A Comparison with Lazy and Eager methods	75
	4.2.1 Backpropagation Neural Network in Learning Phase	75
	4.2.2 Condensed Nearest Neighbor in Testing Phase	78
4.3	Experimental Results for Real Datasets	80
	4.3.1 The XOR Problem	80
	4.3.2 The Haberman Dataset	82
5	CONCLUSION AND FUTURE RECOMMENDATION	
5.1	Conclusion	89
5.2	Future Recommendation	90
	REFERENCES	92
	APPENDICES	97
	BIODATA OF STUDENT	107
	LIST OF PUBLICATIONS	108



LIST OF TABLES

		Page
2.1	A brief detail of some current methods in supervised nonparametric rules for pattern recognition	42
4.1	Accuracy (%) and training time (s) for finding decision boundary with 2,3,and 4 classes artificial data	74
4.2	Training time (s) and Performance goal of Grid Based Classifier and Neural Network	77
4.3	Testing time (s) for Grid based Classifier and Condensed Nearest Neighbor	79
4.4	Truth table for the XOR problem	81
4.5	Truth table for the AND and OR problems	82
4.6	Training time (s) & Training accuracy (%) of Grid Based Classifier and Neural Network	85
4.7	Testing time (ms) & Testing error (%) of Grid Based Classifier and Neural Network by using the ten-fold cross-validation technique.	86
4.8	The best Testing Performance reported for Haberman dataset	87



LIST OF FIGURES

		Page
1.1	Different approaches used for classification	2
1.2	When there is noise, zero misclassification error may not be possible with simple boundary	7
1.3	Linear decision boundary for 2-class classification	9
1.4	Quadratic decision boundary for 2-class classification	9
1.5	Complex decision boundary for 2-class classification	10
1.6	The two separate tasks commonly used within pattern recognition systems	13
2.1	Two-dimensional Parzen-windows made the decision boundaries depend on the window width h	23
2.2	Nearest-neighbor algorithm leads to a partitioning of the input space into voronoi cells	24
2.3	k -nearest-neighbor estimation of two one-dimensional densities	27
2.4	Decision boundaries made by decision trees with perpendicular portions to the feature axes	32
2.5	Three layer neural network with N -dimensional input and M -dimensional output	37
2.6	Decision boundaries made by two-layer network and three-layer network classifier	41
2.7	Complex decision boundary made by an artificial neural network	42
3.1	(a) Training set for the class of data. (b) Example of a hypothesis class	49
3.2	(a) Multiclass features space and training data. (b) Desirable grid which guarantees that each cell contains only one kind of pattern	52



3.3	(a) Finding the decision boundary based on the boundary cells; (b) Eliminate the training data and keep the boundary cells	52
3.4	The flowchart of the main algorithm for the learning stage	53
3.5	(a) Two different patterns with minimum distance. (b) Finding the grid size and plot on the features space	54
3.6	The flowchart of main algorithm for testing stage	56
3.7	Two dimensional training data was generated and plotted in MATLAB	58
3.8	Superpose the grid on training data and determine the decision boundary	60
3.9	The decision boundary was kept in a hash table as a decision function	61
3.10	Three dimensional training data was generated and plotted in MATLAB	62
3.11	Superpose the grid on training data and determine the decision boundary	63
4.1	(a) Data sets 1 to 4 ; (b) Applying the training procedure and (c) Training result	66
4.2	(a) Data sets 5 to 8 ; (b) Applying the training procedure and (c) Training result	67
4.3	(a) Data sets 9 to 12 ; (b) Applying the training procedure and (c) Training result	68
4.4	(a) Data sets 13 to 16 ; (b) Applying the training procedure and (c) Training result	69
4.5	(a) Data sets 17 to 20 ; (b) Applying the training procedure and (c) Training result	70
4.6	(a) Data sets 21 to 24 ; (b) Applying the training procedure and (c) Training result	71



4.7	(a) Data sets 25 to 28 ; (b) Applying the training procedure and (c) Training result	72
4.8	(a) Data sets 29 and 30 ; (b) Applying the training procedure and (c) Training result	73
4.9	A comparison in the training time between the GBC and NN. The numbers above each node in the graph indicates performance goal	78
4.10	(a) Class A and B for XOR problem. (b) Decision boundary by GBC for the XOR problem	81
4.11	Decision boundary by GBC for the (a) AND and (b) OR problems	82
4.12	Haberman dataset is plotted in 3-dimensional space	84
4.13	Decision boundary made by GBC on Haberman dataset	84



LIST OF ABBREVIATIONS

Acronyms

ANN	Artificial Neural Network
GBC	Grid Based Classifier
KNN	K Nearest Neighbor
LM	Levenberg Marquardt
MLP	Multi-layer Perceptron Network
NN	Neural Network
BNN	Backpropagation Neural Net



CHAPTER 1

INTRODUCTION

1.1 Background

Pattern classification is a considerable problem in a various field of scientific disciplines, such as biology, psychology, medical diagnoses, computer vision, artificial intelligence, remote sensing and engineering. Recently, there has been increasing interest in the area of pattern classification specially, the emerging applications which are not only challenging but also computationally more demanding. These applications include data mining, document classification for searching text documents, financial forecasting, organization and retrieval databases of all kind of media, and biometrics (such as face and fingerprints recognition for personal identification based on the various physical attributes). Growing techniques and powerful computing systems, incentive the faster processing on huge data sets, which has also facilitated the use of complicated and diverse methods for pattern classification.

But what is pattern classification? In pattern classification, usually we are given a set of training samples with associated “class labels”. The problem is to design a classifier to predict the class label for any given sample. Pattern classification is one of the two categories of “pattern recognition”. The other category is “clustering”, in which the given samples are not labeled and the problem is to classify the data into groups with features that distinguish one group from another.



1.1.1 Classification

There are various techniques which are applied in the terms of classification. Classification techniques are generally grouped into several categories such as supervised versus unsupervised or parametric versus nonparametric methods (see Fig. 1.1).

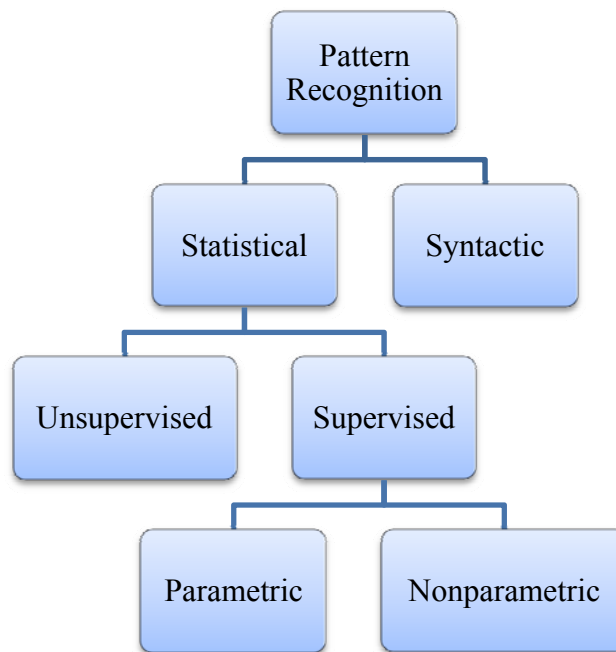


Fig. 1.1 : Different approaches used for classification

In supervised learning, the goal is to make a mapping from the input to an output whose train samples are provided by a supervisor, but in unsupervised learning, the aim is to find the regularities in the input and there is no such supervisor and only input data are provided. Usually there would be a structure to the input space which means some samples happen more often than others, and we try to see as a rule what happens and

what does not. which is sometimes called density estimation, and as a supervised learning the task is to classify samples into one of N different classes based on features of the samples.

1.1.2 Supervised Parametric Methods

In parametric methods, they propose a model valid over the whole input space. When normal density is defined, it means that all examples of the class are drawn from this same density. The supervised parametric method uses the information about data and try to find the probability density function for each class such as maximum likelihood estimation [1,2].

Let w_i for $i = 1, \dots, N$ present the classes. So, if d attributes are considered for each pattern then each pattern is described by a d -dimensional feature vector and then x would present this feature vector. First a set of training pattern is given to the classifier to train. Then the classifier will be allocated a test pattern to a particular class. In the following details of the parametric classifiers are described.

A prior probability for belonging a sample to class w_i denote by $P(w_i)$ and $P(x|w_i)$ also present the class-conditional probability density function which determine the probability function for a feature vector x that belongs to class w_i . The posteriori probability is denoted by $P(w_i|x)$, which is the probability for belonging the pattern to



class w_i for the feature vector x . $P(w_i|x)$ for the feature vector x is determined by the Bayes formula.

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)} \quad (1.1)$$

Where

$$P(x) = \sum_{i=1}^N P(x|w_i)P(w_i) \quad (1.2)$$

This formula is suitable for all probability density functions. Usually the normal density function is used to model the distribution of feature values of a particular class based on the characteristics of the data. The normal density function for d-dimensional data set is given by :

$$P(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right] \quad (1.3)$$

where x is a d elements of feature vector, μ is the d elements of mean vector, Σ denotes the covariance matrix, $|\Sigma|$ and Σ^{-1} respectively denote the determinant and inverse of the covariance matrix.

The μ and Σ of the probability function are calculated from the training data belonging to that class and for each class, the maximum likelihood estimation for μ and Σ are

respectively the mean vector and covariance matrix of the training data of that class. So, by using the Bayes decision rule, we can classify the test sample, illustrated by the feature vector x , which is :

$$\text{decide } w_i \text{ if } P(w_i|x) > P(w_j|x) \quad \forall j \neq i \quad (1.4)$$

The advantage of the parametric method is that it reduces the problem of estimating a probability density function to estimating the value of a small number of parameters. Its disadvantage is that our estimation does not always sustain and it may cause a large error.

1.1.3 Supervised Non-Parametric Methods

There are many classification tasks in which no assumptions can be made about the characterizing parameters. Non-parametric approaches are designed for those tasks. Therefore these algorithms try to find the similar previous instances from the training data set by using a suitable distance measurement from them to find the right output. Different non-parametric methods differ in the way they define similarity and interpolate from the similar training instances. In a parametric model, all of the training samples have an effect on the final estimate, whereas in the non-parametric case, there is no single global model; local models would be estimated as they are needed, affected only by the training instances close by.



Non-parametric methods are also called instance-based learning methods, because they try to store the training samples in a lookup table and using interpolation from these. This implies that all of training instances should be stored and storing all requires memory of $O(N)$. Furthermore, for a given input, look like ones should be found, and finding them requires computation of $O(N)$.

The supervised non-parametric methods selected for classification focus on the data itself and do the classification such as nearest neighbor and the k-nearest neighbor and decision trees methods [1-3]. In such methods there is no need to make assumptions about the probability density function for each class. Their ability would show in their generality. The procedure is same as unimodal normal and the bimodal mixture case and we did not need to make any assumptions about the distributions. If there are enough samples, we will be assured of convergence to an arbitrarily target density. On the other hand, they may need a large number of samples and much more than would be required if we are given the form of the distribution density. However, there are different ways to reduce the data, which leads to reduce requirements for computation time and storage.

1.1.4 Neural Network

There are so many classification methods in use. Apparently among these methods, neural network methods are most widely known. The most important advantages of neural network are the generality of these methods, enable to handle problems with



many vary parameters, and the ability to classify objects correctly even in a very complex distribution of objects. The disadvantages of neural networks are that they are so slow, especially in the learning stage and sometimes in the application stage. Also most of the times, it is very difficult to determine the way of making decision of neural networks.

1.1.5 Noise

Noise is any unwanted inconsistency in the data and because of the noise, the class may be more difficult to learn and zero error may be impracticable with a simple hypothesis class (see figure 1.2).

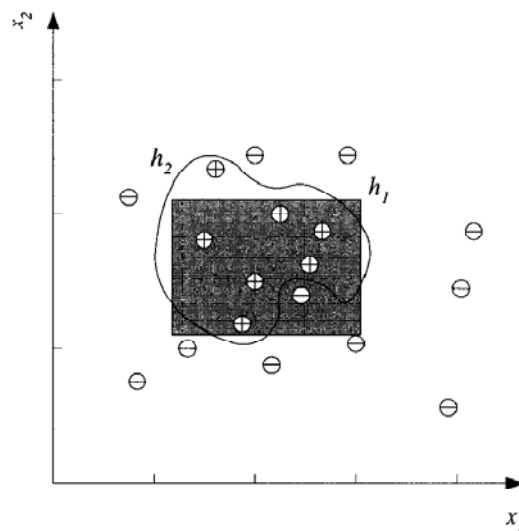


Fig. 1.2 : When there is noise, zero misclassification error may not be possible with simple boundary.

There are several explanations of noise: