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 digunapakai 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 & Masoomeh 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

LIST OF TABLES

LIST OF FIGURES

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).

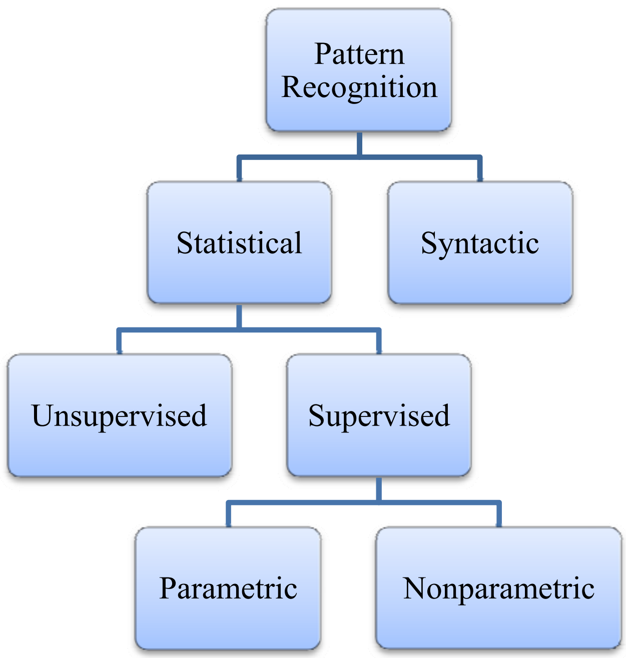

Fig. 1.1 : Different approaches used for classification

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 drowning 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 $i = 1, \ldots ,N$ 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.

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

Where

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

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 :

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

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 :

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

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).

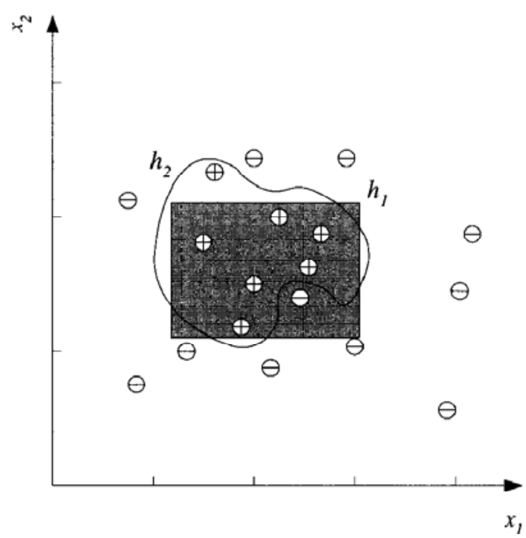

Fig. 1.2 : When there is noise, zero misclassification error may not be possible with simple boundary.

There are several explanations of noise: