UNIVERSITI PUTRA MALAYSIA

A HYBRID ROUGH SETS K-MEANS VECTOR QUANTIZATION MODEL FOR NEURAL NETWORKS BASED ARABIC SPEECH RECOGNITION

ELSADIG AHMED MOHAMED BABIKER

A HYBRID ROUGH SETS K-MEANS VECTOR QUANTIZATION MODEL FOR A NEURAL NETWORKS BASED ARABIC SPEECH RECOGNITION

By

ELSADIG AHMED MOHAMED BABIKER

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

September 2002

DEDICATION

To:

My parents, brother and sisters,
My wife,
My children Husam & Aala

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy

A HYBRID ROUGH SETS K-MEANS VECTOR QUANTIZATION MODEL FOR A NEURAL NETWORK BASED ARABIC SPEECH RECOGNITION

By

ELSADIG AHMED MOHAMED BABIKER

September 2002

Chairman: Abd Rahman Ramli, Ph.D.

Faculty: Engineering

Speech is a natural, convenient and rapid means of human communication. The ability to respond to spoken language is of special importance in computer application wherein the user cannot use his/her limbs in a proper way, and may be useful in office automation systems. It can help in developing control systems for many applications such as in telephone assistance systems.

Rough sets theory represents a mathematical approach to vagueness and uncertainty. Data analysis, data reduction, approximate classification, machine learning, and discovery of pattern in data are functions performed by a rough sets analysis. It was one of the first non-statistical methodologies of data analysis. It extends classical set theory by incorporating into the set model the notion of classification as indiscernibility relation.

In previous work rough sets approach application to the field of speech recognition was limited to the pattern matching stage. That is, to use training speech patterns to generate classification rules that can be used later to classify input words patterns.

In this thesis rough sets approach was used in the preprocessing stages, namely in the vector quantization operation in which feature vectors are quantized or classified to a finite set of codebook classes. Classification rules were generated from training feature vectors set, and a modified form of the standard voter classification algorithm, that use the rough sets generated rules, was applied.

A vector quantization model that incorporate rough sets attribute reduction and rules generation with a modified version of the K-means clustering algorithm was developed, implemented and tested as a part of a speech recognition framework, in which the Learning Vector Quantization (LVQ) neural network model was used in the pattern matching stage.

In addition to the Arabic speech data that used in the original experiments, for both speaker dependant and speaker independent tests, more verification experiments were conducted using the TI20 speech data.

The rough sets vector quantization model proved its usefulness in the speech recognition framework, however it can be extended to different applications that involve large amounts of data such as speaker verification.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah.

MODEL HIBRID SET KASAR K-MEANS PENGKUANTUMAN VEKTOR UNTUK RANGKAIAN NEURAL YANG BERASASKAN PENGECAMAN PERCAKAPAN ARAB

Oleh

ELSADIG AHMED MOHAMED BABIKER

September 2002

Pengerusi: Abd Rahman Ramli, Ph.D.

Fakulti : Kejuruteraan

Percakapan adalah semulajadi, mudah dan cepat untuk perhubungan di antara manusia. Kebolehan untuk bertindak balas ke atas percakapan adalah penting untuk aplikasi komputer sekiranya pengguna tersebut mempunyai kecacatan tangan atau di pejabat di mana sistem aplikasi tersebut dikawal secara automatik. Kebolehan ini juga dapat membantu di dalam membangunkan sistem kawalan untuk pelbagai aplikasi lain seperti sistem bantuan berasaskan telefon.

Teori set-set kasar mewakili suatu pendekatan bermatematik untuk ketakjelasan dan ketaktentuan. Fungsi untuk analisis data, pengurangan data, pengelasan penghampiran, pembelajaran mesin dan penemuan corak data adalah beberapa fungsi dijalankan oleh analisis set set kasar. Ia adalah di antara kaedah bukan statistik yang pertama digunakan untuk menganalisis data. Ia memperkembangkan teori set klasik dengan memasukkan hubungan tidak boleh dibezakan sebagai suatu pengelasan.

Sebelum ini, penggunaan set-set kasar di dalam bidang pengecaman percakapan adalah terhad kepada peringkat pemadanan corak. Kaedah tersebut menggunakan corak percakapan sebagai latihan untuk menjana peraturan pengelasan yang boleh digunakan untuk mengelaskan corak masukan perkataan.

Di dalam tesis ini, pendekatan set-set kasar digunakan di dalam peringkat pra-pemprosesan, iaitu di dalam peringkat operasi pengkuantuman vector di mana ciri-ciri setiap vector dikuantumkan atau dikelaskan kepada set terhingga kelas-kelas buku-kod. Peraturan pengelasan telah dijanakan daripada latihan ciri-ciri set vector-vektor dan algoritma pengelasan piawai pengundi yang telah diubahsuai. Peraturan-peraturan yang digunakan adalah hasil daripada penjanaan oleh set-set kasar.

Model pengkuantuman vektor yang mengandungi pengurangan sifat set-set kasar dan penjanaan peraturan-peraturan dengan menggunakan versi algoritma pengumpulan K-purata yang diubahsuai telah dibangunkan, dilaksanakan dan diuji. Model ini adalah sebahagian daripada kerangka kerja pengecaman percakapan di mana model rangkaian neural (LVQ) telah digunakan pada peringkat pemadanan corak.

Sebagai tambahan kepada percakapan bahasa Arab yang digunakan di dalam ujikaji awal, pengesahan ujikaji juga dijalankan menggunakan data percakapan TI20 untuk ke dua-dua ujian bersandar penutur dan tiada bersandar penutur.

Model set-set kasar pengkuantuman vektor membuktikan kegunaannya di dalam kerangka kerja pengecaman percakapan. Walau bagaimanapun ia boleh

diperkembangkan kepada aplikasi lain yang mempunyai data yang banyak seperti pengesahan penutur.

ACKNOWLEDGEMENTS

I am deeply thankful to my supervisor Dr. Abd Rahman bin Ramli for his immense support, continuous encouragement, and valuable guidance throughout the course of my research.

I am very much grateful to Assoc. Prof. Dr. Bachok M. Taib and Dr. Zainul Abidin Hasan for their constructive comments, guidance and support.

I am much deeply grateful to my parents; from them I learned every thing that I need in my life and thereafter.

The rest of my family has provided support and, more important still, motivation and great relief. I cannot forget here the roll of my brother Dr. Khalid, my sisters and Salah M. A.

I extend my deepest appreciation to my wife, to whom I owe so much, she manage to understand the work presented in this thesis before any body and helped by providing useful suggestions in all stages. My children Husam and Aala contributed to this work in their own way.

I am very thankful to Mr. Saiful who helped in producing the Malay version of the abstract.

Special thanks to the staff of the School of Graduate Studies-UPM for their unlimited patience and high quality service.

Last but not least, I am very much thankful to all my friend who helped me in one form or another.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ANNs	Artificial Neural Networks
ARPA	Advanced Research Projects Agency
ASR	Automatic Speech Recognition
CB	Code Book
CW	Code Word
DARPA	Defense Advanced Research Projects Agency
DSP	Digital Signal Processing
DTW	Dynamic Time Warping
HMM	Hidden Markov Model
HMMs	Hidden Markov Models
IS	Information System
KSOM	Kohonen Self-Organizing Map
LHS	Left Hand Side
LPC	Linear Predictive Coding
LVQ	Learning Vector Quantization
MEC	Minimum Error Classification
MLP	MultiLayer Perceptron
MSE	Mean Square Error
NIST	National Institute of Standards and Technology
NNET	Neural Network
NNETs	Neural Networks
RHS	Right Hand Side
RIFF	Resource Interchange File Format
RNN	Recurrent Neural Networks
RS	Rough Sets
RSKMVQM	Rough Sets – K-Means Vector Quantization Model
SLIs	Spoken Language Interfaces
SOFM	Self-Organizing Feature Maps
SOM	Self-Organizing Maps
SPHERE	SPeech HEader REsources Package
SR	Speech Recognition
SRS	Speech Recognition System
SUS	Speech Understanding Systems
TDNN	Time-Delay Neural Network
TI	Texas Instruments
TI-20	Texas Instruments 20 words speech corpora
TI-46	Texas Instruments 46 words speech corpora
VQ	Vector Quantization

CHAPTER 1

INTRODUCTION

1.1 Overview

Research in the area of automatic speech recognition started before the early days of computers. Now it can be considered as one of the hottest areas of research in computer science, see (Gold and Morgan 2000, Pierce 1969 and Potter *et al* 1947).

We are currently in the midst of a revolution in communications that promises to provide unlimited everywhere access to multimedia communication services. In order to succeed, this revolution demands faultless, easy-to-use, high quality interfaces to support broadband communication between people and machines. Spoken language interfaces (SLIs) are essential for making this vision a reality. Motivations for this potential applications of SLIs and the underlying technologies are key areas for future research in both spoken language processing and human computer interfaces (Candace *et al* 1997).

From a commercial point of view, speech recognition is a technology with a large market potential. It can be of use for voice-activated transcription, telephone assistance, office automation systems, and many other man-machine interface tasks.

Speech recognition is the ability on part of a computer to accept speech signals as input, and produce as output a sequence of words that corresponds to the spoken input. In other words, automatic speech recognition (ASR) is the process by which a computer maps an acoustic speech signal to text (Rabiner and Juang 1993, Linggard *et al* 1992 and Nakagawa *et al* 1995).

Rough sets theory represents a mathematical approach to vagueness and uncertainty, data analysis, data reduction, approximate classification, machine learning, and discovery of pattern in data are functions performed by a rough sets analysis. It was one of the non-statistical methodologies of data analysis. It extends classical set theory by incorporating into the set model the notion of classification as indiscernibility relation (Slowinski 1992 and Pawlak 1991).

Using rough sets approach in the field of speech recognition was limited in previous work to the pattern matching stage. That is, to use training speech patterns to generate classification rules that can be used later to classify input words patterns (Czyzewski and Kaczmarek 1995, Kostek 1995, Czyzewski and Kaczmarek 1998 and Kostek 1998).

In this work a new vector quantization model that combines a modified form of the K-means clustering algorithm and the rough sets approach attribute reduction, rule generation and classification algorithms was developed and tested. Classification rules were generated from training feature vectors set, and a modified form of the

standard voter classification algorithm, that use the rough sets generated rules, was applied.

1.2 Objectives and Scope of This Work

The main goal for any speech recognition research work is to minimize the amount of data that the speech signal carries in each stage of the speech recognition system (SRS) without loosing any part of the information that the speech signal represents. Information here refers to a higher level than such a pure data; that is information on phonemes and utterances. Researchers usually focus their work on improving the performance in one stage of the recognition process.

Vector Quantization (VQ) is almost considered as a data compression technique by digital signal processing (DSP) researchers. As a matter of fact, a lot of work has been done in this field to improve the quality of speech and image processors. Vector quantization has been applied more and more to reduce the complexity of problems like pattern recognition. In speech recognition, vector quantization used to preprocess speech signal prior to the pattern matching stage. Discrete systems are welcome in real-time implementations since they are less CPU consuming than continuous systems. Better results were obtained in speech recognition due to improvements in vector quantization techniques, see for example Fontaine *et al* (1994). Although VQ is ignored sometimes in speech recognition implementations, It has been shown that VQ methods and their parameters have an important influence

on recognition rates. Therefore, some experiments have been done on vector quantization in the framework of a complete speech recognition system. The same recognition algorithms and the same feature vectors have been used to make the comparisons. Many researchers showed the influence of vector quantization on speech recognition system's performance (Fontaine *et al* 1994 and Huang and Kuh 1992).

Methods used for VQ are mainly numerical; they work on continuous real valued attributes, this implies that any small difference in attribute values (such as energy or frequency) for two vectors may result in classifying the two vectors in two classes or clusters, while discrete values methods discard small variations in attribute values. Small variations in attribute values may be a result of external noise in the recording environment or may also be a result of the natural variation in pronouncing an utterance more than once even by the same speaker. Rough sets (RS) approach, which is a relatively new approach to data analysis under uncertainty works mainly on discrete valued attributes. Under rough sets; there are two main training steps in data analysis: first reduction of attributes, and second rule generation. Generated rule can then be used to classify new objects; previous research work on applying rough sets to speech recognition analysis considered complete utterances as information system's objects ignoring the RS data reduction power. That is to apply RS to the pattern matching stage rather than vector quantization stage.

Additionally, most of the research work in the area of speech recognition was oriented towards English language, less work was done for other languages such as

French, Spanish and Mandarin, and very little amount of work was done for Arabic language which has its own special features and importance.

The objectives of this work are:

1) To develop a vector quantization model that employs the rough sets attribute reduction, rule generation, and objects classification algorithms; this model combines rough sets algorithms with the K-means clustering algorithm.
2) To incorporate the developed model in an Arabic speech recognition system that uses neural networks for pattern matching.
3) To compare recognition rates results obtained from the developed model with the pure K-means algorithm vector quantization.
4) To test and verify the performance of the developed model using English vocabularies.

1.3 Organization of the Thesis

After this introduction this thesis is formed of four section; these are:

Literature review; this mainly in chapter two,

Description of the methods used in this work; this is distributed in three chapters; chapter three, four and five.

In chapters six and seven; A description for the rough sets vector quantizer and Arabic speech recognizer, with a presentation for the experiments and their results with a discussion.

Conclusions and suggestions for further work; this is the subject of chapter eight.

The details of the thesis chapter is as follows:

- Chapter One; Introduction: A quick overview of the field of speech recognition and rough sets, followed by a statement of objectives and scope of this work, and then the organization of the thesis.
- Chapter Two; Literature review: Literature on speech recognition and related topics and methods, rough sets, neural networks and their application was reviewed in this chapter.
- Chapter Three; Automatic Speech Recognition Fundamentals; the basic ideas behind speech recognition: different approaches to speech recognition, speech systems classifications, with a brief idea on Arabic and English phonetic. The structure of a typical speech recognition system with some details for the system components is presented as well as the methods and techniques used in each component.
- Chapter Four; Rough sets theory and applications: Fundamentals of the theory that proposed by Pawlak in the early 1980's, concepts of information systems and their reductions, and rule generation were presented followed by a applications of rough sets.
- Chapter Five; Neural networks and applications: In addition to neural networks basic concepts, Self Organizing Maps and Learning Vector Quantization neural network models were discussed here, followed by a brief description for the application of neural networks to speech recognition.