UNIVERSITI PUTRA MALAYSIA

PARALLEL IMPLEMENTATION ON IMPROVED ERROR SIGNAL OF BACKPROPAGATION ALGORITHM

TEH NORANIS BT MOHO ARIS

FSKTM 2001 10

PARALLEL IMPLEMENTATION ON IMPROVED ERROR SIGNAL OF BACKPROPAGATION ALGORITHM

By

TEH NORANIS BT MOHD ARIS

Thesis Submitted in Fulfilment of the Requirements for the Degree of Master of Science in the Faculty of Science Universiti Putra Malaysia

May 2001

Dedicated to my husband, Shahrin Azuan,

my daughter, Nisa Syakirah,

my son, Muhammad Rafiq,

my parents and family.

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

PARALLEL IMPLEMENTATION ON IMPROVED ERROR SIGNAL OF BACKPROPAGATION ALGORITHM

By

TEH NORANIS BT MOHD ARIS

May 2001

Chairman: Associate Professor Md. Yazid Mohd Saman, Ph.D.

Faculty : Computer Science and Information Technology

The research work presented in this thesis is a continuation of Shamsuddin's work regarding proposed error signal for the backpropagation (BP) algorithm. The main focus is to parallelise Shamsuddin's work in order to improve the speedup of the BP algorithm. The experiments are implemented using the Sequent Symmetry SE30 parallel machine. The BP algorithm uses the data partitioning method with columnwise block striped and the batch mode weight updating strategy. Twenty-six patterns consisting of uppercase letters from 'A' to 'Z' are tested in the experiments. Two main factors taken into consideration in this, experiments are the execution time and speedup and the recognition rates.

Shamsuddin's proposed BP parallel version, is compared with the sequential version. Experimental results shows that the execution time of the parallel version is much less than the execution time of the sequential version. The parallel version produces a good speedup as the number of processors, are increased due to the value that is near the ideal value.

Experiments for testing the recognition rates involves the twenty-six trained sample data with perfect pattern and untrained sample data with 10% corrupted pattern. The recognition rates results show 100% accuracy for the trained and untrained data using the standard BP and Shamsuddin's proposed BP running sequentially.

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

PERLAKSANAAN SECARA SELARI ALGORITMA RAMBATAN BALIK ISYARAT RALAT YANG TELAH DIPERBAIKI

Oleh

TEH NORANIS BT MOHD ARIS

Mei 2001

Pengerusi: Profesor Madya Md. Yazid Mohd Saman, Ph.D.

Faculti : Sains Komputer dan Teknologi Maklumat

Kerja-kerja yang dibentangkan dalam tesis ini adalah sambungan kepada kerja Shamsuddin yang berkaitan dengan kaedah isyarat ralat bagi algoritma rambatan balik (BP). Fokus utama adalah menjalankan kerja Shamsuddin secara selari untuk memperbaiki kelajuan algoritma BP. Pengujianan dilaksanakan menggunakan mesin selari Sequent Symmetry SE30. Algoritma BP ini menggunakan kaedah pembahagian data dengan strategi jalur blok berdasarkan lajur dan mod kelompok pengemaskinian pemberat. Dua puluh enam paten yang terdiri daripada huruf besar dari ‘A’ ke ‘Z’ diuji dalam experimen. Dua faktor utama yang ditekankan dalam eksperimen ini ialah masa larian dan kelajuan dan kadar pengecaman.

Versi selari cadangan BP Shamsuddin's dibandingkan dengan versi berturutan. Keputusan eksperimen menunjukkan masa larian bagi versi selari adalah jauh lebih kurang berbanding masa larian bagi versi berturutan. Versi selari menghasilkan kelajuan yang baik apabila bilangan pemproses ditambah kerana nilai kelajuan adalah berhampiran dengan nilai kelajuan mengikut teori.

Eksperimen bagi menguji kadar pengecaman melibatkan dua puluh enam data sampel dengan paten lengkap yang dilatih dan data sampel dengan paten yang dirosakkan sebanyak 10% yang tidak dilatih. Kadar pengecaman menunjukkan ketepatan pengecaman 100% bagi data yang dilatih dan tidak dilatih menggunakan BP piawai dan BP cadangan Shamsuddin yang dilarikan secara berturutan.

ACKNOWLEDGEMENTS

I would like to express my most gratefulness and faithful appreciation to the Chairman of my Supervisory Committee, Associate Professor Dr. Md. Yazid Bin Mohd Saman for his precious guidance, motivation and advice throughout my studies. I would also like to acknowledge the members of my Supervisory Committee, Dr. Md. Nasir Bin Sulaiman of the Department of Computer Science and Dr. Mohamed Bin Othman of the Department of Communication Technology and Network. Their suggestions and comments are valuable to the completion of this thesis.

Special thanks to the Dean of Faculty of Computer Science and Information Technology, Dr. Abdul Azim Bin Abd Ghani, Deputy Dean of Faculty of Computer Science and Information Technology, Dr. Ramlan Bin Mahmod and Head of Computer Science Department, Associate Professor Dr. Ali Bin Mamat. I also like to thank the academic and administrative staff of Faculty of Computer Science and Information Technology, Universiti Putra Malaysia for their assistance.

Thank you to my helpful husband for his encouragement, support and motivation throughout my studies. Special thanks for the help with my research work, to my friends who are also researchers in this field especially Dr. Rozita Bt Johari, Razali Bin Yaacob, Ummu Salmah Bt Mohd Hussin, Norhayati Bt Abdullah, Azuraliza Bt Abu Bakar and Lee Lai Soon.

This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfilment of the requirement for the degree of Master of Science.

AINI IDERIS, Ph.D.,
Professor,
Dean of Graduate School,
Universiti Putra Malaysia.

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADALINE	-	Adaptive Linear Neuron
ART	-	Adaptive Resonance Theory
BAM	-	Bidirectional Associative Memory
BP	-	Backpropagation
BSB	-	Brain-State-in-a-Box
CPN	-	Counterpropagation Network
CPU	-	Central Processing Unit
ENIAC	-	Electronic Numerical Integrator and Calcalator
IeBP	-	Improved error of BackPropagation algorithm
I/O	-	Input/Output
MADALINES	-	Multiple ADALINEs
MIMD	-	Multiple Instruction Stream, Multiple Data Stream
MISD	-	Multiple Instruction Stream, Single Data Stream
MSE	-	Mean Squared Error
NN	-	Neural Network
SIMD	-	Single Instruction Stream, Multiple Data Stream
SISD	-	Single Instruction Stream, Single Data Stream
VLSI	-	Very Large Scale Integration

CHAPTER I

INTRODUCTION

Background

There is a high demand of computational speed for a great number of areas such as numerical modeling and simulation of scientific and engineering problems (Wilkinson and Allen, 1999). These problems require huge repetitive calculations on large volumes of data to give valid results. The computations must be fast and completed within a time period. Solving these problems using parallel computers is the answer. Parallel computers consist of several processors running concurrently. The execution speed is much faster compared to a computer with a single processor.

Artificial neural networks or referred to, as neural network (NN) is one of the artificial intelligence areas which has a close connection with parallel processing. NN attempts to imitate the computational power of the human brain. The human brain characteristic is a highly complex, nonlinear, and parallel processing system (Haykin, 1999). It has the powerful capability to perform certain computations such as pattern recognition, perception and motor control many times faster than the fastest digital computer available today.

A NN model consists of a massive interconnection of simple computing cells called nodes or neurons. Each node is connected to other nodes by directed communication links. Each node is also provided with an activation level and an associated weight. The activation level produces the output of the node. The weights contain fundamental information concerning the problem being solved by the NN. The weights are adjusted in a step by step procedure called the training process. The training process is repeated until the NN reaches a stage where it is well trained.

NN training process is time consuming. Therefore NN simulation requires computational speed in order to reduce the execution time of the training process, which involves repetitive calculations. The backpropagation (BP) algorithm (Rumelhart *et al.*, 1986) is one of the most popular NN algorithms. It has been used in a large number of applications (Shekhar and Amin, 1992, Dutta and Shekhar, 1988, White, 1988 Sejnowski and Rosenberg, 1986). Much research has been performed to speed up the BP training process. Two approaches used are improving the BP algorithm or implementing the BP using parallel machines (Mangasarian and Solodov, 1994).

Shamsuddin (2000) has proposed an error signal for the BP NN algorithm, hereafter will be called IeBP (Improved error of BackPropagation algorithm). A modified error function has been generated to increase the convergence rates of the BP training, replaced by the Mean Squared Error (MSE) used in standard BP. From the experimental results, the IeBP also proved that the epoch size of the modified BP is less than the epoch size of the standard BP. Therefore, the execution time of the IeBP is faster than the standard BP. The experiments are carried out using a sequential computer.

The usage of parallel computers is becoming popular after 40 years of complete focus on sequential computers (Lester, 1993). The growth of Very Large Scale Integrated (VLSI) processor had produced high speed computers which operate in parallel. The approach used is to assemble together large numbers of VLSI chips in one computer.

The massively parallel characteristics of NN make it very well suited to be implemented using parallel processors. In addition, much research had focused on parallel implementations of NN (Ammar *et al.*, 1998).

Problem Statement

As mentioned earlier, the IeBP introduced by Shamsuddin (2000) has been implemented on a sequential computer. However, as the input data sets become large, the execution time of the IeBP becomes slow. Therefore, the execution time can be improved by implementing the IeBP on a parallel processor. Thus, in this research, applying parallel processing to the IeBP is the main focus.

Objectives of the Research

The objective of this research is to combine the usage of parallel processing and IeBP to produce a much faster BP training algorithm. The detailed objectives of this research are as follows:

i. To parallelise the standard BP and IeBP controlled by the number of cycles using the columnwise block striping parallel method on the Sequent Symmetry SE30 shared memory machine for pattern recognition application.

ii. Compare the parallel version of the standard BP and the IeBP in terms of speedup.

iii. Test recognition rates of IeBP.

Scope of the Research

The scope of the research is limited to the input data consisting of twenty-six uppercase patterns from ‘A’ to ‘Z’ with size four hundred binary inputs. This input data is used as the trained data set. The trained data set is used during the training process and the recall back process. Another set of data, which is the same as the input data mentioned but 10% corrupted is used as untrained data. The untrained data is used during the recall back process.

Methodology

The standard BP and the IeBP are developed using the C programming language. The parallel programming method applied is data partitioning. The data partitioning schemes used is columnwise block striping. Columnwise block striping method is used due to the number of input columns (four hundred), which is greater than the number of row patterns (twenty-six) resulting in less execution time. Another method, which is the rowwise cyclic striping is adopted by Sanossian (1992) and Sulaiman and Evans (1996). They used the rowwise cyclic striping method because the number of input columns (fourty) is less than the number of row patterns (fifty). In this research, the weight updating strategy applied is batch mode strategy and training set parallelism. Batch mode strategy and training set parallelism is used because experiments from previous research (Sanossian, 1992, Sulaiman and Evans, 1996 and Ammar and Miao, 2000) proved that the speedup produced is much better than other types of parallelism.

Thesis Organization Structure

This thesis consists of five chapters, including the introduction chapter, which explains in general about the background of parallel processing, NN and the IeBP. The problem statement discussed suggestion to improve IeBP algorithm using parallel processor. The objectives of the research, scope of the research and methodology are also discussed in the introduction chapter.

In Chapter II, explanation on NN history, NN capabilities, NN learning, NN architectures, activation functions, standard BP and IeBP are given. This chapter also briefs the computing history, parallel paradigms, Sequent Symmetry SE30 Architecture, elements of parallel programming on Sequent SE30, performance measurements, Amdahl's Law and Gustafson's Law. In addition, related research work, are also explained in this chapter.

Chapter III describes the parallel paradigms for NN, the Sequential BP and the general workflow of the whole system. In addition, this chapter includes detailed design framework of parallel BP.

Implementation issues on the parallel techniques are discussed in Chapter IV. It also presents the experimental results. Finally, in Chapter V the thesis, concludes the research work and suggest recommendations for further work.

CHAPTER II

BACKGROUND OF PARALLEL NEURAL NETWORK

Introduction

NN has been used to solve the real-world application problems, such as pattern recognition, vision and speech recognition. In this research, pattern recognition application will be applied. The ability of NN to adapt is very important in the pattern recognition area. In addition, the massively parallel nature of NN makes it very well suited to be implemented using parallel processors.

This chapter includes explanations on NN, the standard BP, parallel computing and the Sequent Symmetry SE30 parallel machine, which is used as a platform in this research. The last section will stress on the previous research work on leBP and parallel NN.

Neural Network

NN can be described from two viewpoints, artificial NN and biological NN (Fausett, 1994). Artificial NN processing is carried out in simple processing elements called neurons, nodes, units or cells as shown in Figure 1. Each neuron is connected to each other with an associated