ABSTRACT

Hybridization of Self Organizing Map (SOM) and Particle Swarm Optimization (PSO) is commonly implemented in clustering domain due to its capabilities in handling complex data characteristics. However, some of these hybrid architectures have weaknesses such as slow convergence time; always being trapped in the local minima and others. This study proposes a hybridization method by improving the Self Organizing Map (SOM) Lattice Structure with Particle Swarm Optimization (ESOMPSO) for solving classification problems. The enhancement of SOM lattice structure is implemented by introducing a new hexagon formulation for better mapping quality in data classification and labeling. The improvement of the SOM lattice structure using the proposed Enhanced SOM is implemented by optimizing the weights using PSO to obtain better output quality. The process is done in two stages: the first stage is conducted by training the weights using the Enhanced SOM, and the second stage is implemented by optimizing these weights with the PSO. The proposed method has been tested on various standard datasets. The comparisons are done on standard SOM, Enhanced SOM (ESOM), SOMPSO and ESOMPSO using various distance measurements. The performance of the proposed method is validated using classification accuracy and quantization error. The experiments have shown that ESOMPSO yields promising result with better average accuracy and quantization errors.

ABSTRAK

Penghibridan kaedah Peta Swa Organisasi (PSwaO) dengan Pengoptima Partikel Berkumpulan (PPB) lazim dilaksanakan dalam bidang pengelompokan. Kelaziman ini berlaku disebabkan oleh keupayaan teknik penghibridan dalam mengendalikan ciri-ciri data yang rumit. Walau bagaimanapun, terdapat beberapa kelemahan dalam senibina teknik penghibridan. Oleh yang demikian, kajian ini mencadangkan teknik penghibridan dengan menambahbaik struktur kekisi PSwaO bersama-sama dengan kaedah PPB bagi menyelesaikan masalah pengelasan. Pembaikan struktur kekisi PSwaO dilaksanakan dengan memperkenalkan rumus kekisi heksagon yang baru bagi meningkatkan kualiti pemetaan. Ini bertujuan bagi mendapatkan keputusan yang baik bagi pengelasan dan pelabelan data. Sorotan pengemaskinian pemberat daripada PSwaO dioptimumkan selanjutnya dengan menggunakan PPB bagi mendapatkan kualiti output yang baik. Proses pelaksanaan di atas boleh dibahagikan kepada dua fasa. Fasa pertama melibatkan latihan pemberat menggunakan teknik pembaikan PSwaO. Manakala fasa kedua mengoptimakan pemberat tersebut dengan menggunakan teknik PPB. Teknik PSwaO diuji menggunakan pelbagai data piawaian yang lazim digunapakai oleh penyelidik setara. Perbandingan dapatan kajian terhadap PSwaO dilaksanakan dengan menggunakan pelbagai dimensi jarak terhadap PSwaO tradisi, pembaikan PSwaO, penghibridan PSwaO tradisi dan PBB serta pembaikan PSwaO dengan PPB. Tahap prestasi kaedah cadangan PSwaO diukur menggunakan rumus ketepatan pengelasan dan ralat pengkuantuman. Hasil dapatan kajian menunjukkan bahawa teknik yang diperkenalkan iaitu PSwaO memberi keputusan yang menyakinkan dengan purata ketepatan dan ralat pengkuantuman yang baik.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

In classification process, normally, large classes of objects are separated into smaller classes. This approach can be very complicated due to the challenge in identifying the criteria especially for procedures that involve complex data structures. In this scenario, practically, Machine Learning (ML) techniques will be used and introduced by many researchers as alternative solutions to solve the above problems. Among ML methods, Artificial Neural Network (ANN), Fuzzy Set, Genetic Algorithm (GA), Swarm Intelligence (SI) and rough set are commonly used by the researchers.

However, the most popular ML method widely used by the practitioners is ANN (Negnevitsky, 2005). ANN is also known as neurocomputers or connectionist networks or parallel distributed processors (Haykin, 1999). ANN mimics the biological characteristics of human brain. It involves artificial neurons that can portray the complex universal behavior. ANN learning characteristics is determined by the links between neurons and essentials parameters. In ANN model, simple neurons are connected together to form series of connected network. While a neural network does not have to be adaptive, its advantages arise with proper algorithms to update the weights of the connections to produce a desired output.

ANN and evolutionary computation methodologies have been proven effective in solving certain classes of problems. For example, neural networks are very good at mapping input vectors to outputs and evolutionary algorithms are very good at optimization (Kennedy, 2001). Evolutionary computation is based on population of optimization techniques such as Evolutionary Algorithm (EA) and Swarm Intelligence (SI). Genetic Algorithm (GA) is one of the common techniques used in EA, and it is inspired by biological evolution such as inheritance, mutation, selection and crossover. On the other hand, SI methods such as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) are motivated by flock of bird, swarm of bees, ant colony and school of fish.

The searching implementation with evolutionary method such as ANN learning may overcome the gradient based handicaps. But the convergence is in general much slower, since these are general purpose methods (Branke, 1995). Kennedy and Eberhart proposed nonlinear optimization technique which is very simple, so-called PSO. PSO requires few computational costs (Kennedy & Eberhart, 1995). The authors argued that PSO could train Feedforward Neural Network (FNN) with a performance similar to the Backpropagation (BP) method, for the XOR and Iris benchmarks. Since then, several researchers have adopted PSO for FNN learning (Bergh, 2000). However, most of the studies focus on the hybridization of PSO and FNN. Few studies have been conducted on the hybridization of PSO with Self Organizing Map (SOM) to solve complex problems as given in the next section.

1.2 Problem Background

Early studies have shown that the combination of PSO-SOM approach was first introduced by Shi and Eberhart with modified particle swarm optimizer (Shi and Eberhart, 1998). Subsequently, Xiao *et al.,* (2003;2004) used hybrid SOM-PSO approach to produce better clustering of gene datasets. They used SOM learning and PSO to optimize the weights of SOM. However, the merit for combination of SOM-PSO without conscience factor was poor than SOM alone. This is due to the used of conscience factor that is valuable as a competitive learning technique that reduces the number of epochs necessary to produce a robust solution.

In 2006, O'Neill and Brabazon adopted PSO as unsupervised SOM algorithm. The authors has suggested using different distance metric in calculating the distance between input vectors and each member of the swarm to produce competitive result for data classification. However, in this study, types of SOM lattice structure are not considered.

Chandramouli (2007) used SOM and PSO for image classifier. He stated that SOM is dominated in image classification problems. However, the problem comes in generating image classes which provide concise visualization of the image dataset. Therefore, the author used dual layer of SOM structure and PSO to optimize the weights of SOM neurons.

In 2008, Forkan and Shamsuddin, introduces a new method for surface reconstruction based on hybridization of SOM and PSO. They used growing grid structure in Kohonen network to learn the sample data through mapping grid and PSO to probe the optimum fitting points on the surface. In this study, the proposed Kohonen network is a 3D rectangular map and being enhanced using growing grid method. However, this study doesn't focus on the lattice structure of Kohonen network.

Furthermore, Sharma and Omlin (2009) utilize a U-matrix of SOM to determine cluster boundaries using PSO algorithm. The authors compared the results with other clustering techniques such as k-means and Hierarchical clustering. However this study doesn't focus on the structure of SOM architecture.

Recently, Ozift *et al.,* (2009) proposed PSO in the optimization of SOM algorithm to reduce the training time without loss of quality in clustering. They stated that the size of lattice is related to the clustering quality of SOM. This optimization technique has successfully reducing the numbers of nodes that finds the BMU for a particular input. By having larger grid size in SOM will invite higher training time. Furthermore, the larger the lattice size, the more nodes should be considered for BMU calculation. This causes higher operating cost for the algorithm.

Due to limitations of the previous studies on focusing the improvement of SOM lattice structure, hence, this study will enhance SOM lattice structure by improving the quality of data classification and labeling with improved hexagonal lattice area proposed by Bariah (2007). Particle Swarm Optimization (PSO) is developed to optimize SOMs' training weights accordingly. The hybridization of SOM-PSO architecture, so-called Enhanced SOM with Particle Swarm Optimization (ESOMPSO) is proposed with improvement on the lattice structure for better classification. The performance of the proposed ESOMPSO is validated based on the classification accuracy and Quantization Errors (QE).

1.3 Problem Statement

This research is focus on enhancement of SOM with PSO for classification problems. The improvement of SOM's lattice area and PSO as the optimization technique can improve the generalization of SOM learning characteristics. Hence, the primary main research question for this study is stated as:

Could the Particle Swarm Optimization enhance the learning capability of Self Organizing Map with Improved Lattice Structure?

This lead to the secondary research questions as below:

i) *How to enhance Self Organizing Map (SOM) with Particle Swarm Optimization (PSO)?*
ii) *How efficient is the improved lattice structure for SOM with different distance measurements?*
iii) *How efficient is the enhanced SOM with PSO for classification problems?*
iv) *How to evaluate the performance of the proposed methods?*

1.4 Research Aim

The aim of this study is to enhance Self-Organizing Map (SOM) learning with improved lattice structure and Particle Swarm Optimization (PSO). This proposed method is evaluated, validated and analyzed in terms of QE and classification accuracy. Different types of distance measurements are implemented on the proposed method to real-valued dataset for classification problems.

1.5 Objectives of the Study

To achieve the aim of the study, few objectives have been identified as below:

1. To develop SOM network with improved hexagonal lattice structure (ESOM).
2. To optimize ESOM with Particle Swarm Optimization (ESOMPSO).
3. To design and evaluate standard SOM, ESOM, standard SOM with PSO (SOMPSO) and ESOMPSO using various distance measurements.
4. To compare the performance of standard SOM, ESOM, SOMPSO and ESOMPSO using standard datasets in terms of quantization errors and accuracy.

1.6 Scope of the Study

Since this study is focusing on the enhancement ESOM with improve lattice structure, the scope are bounded as follow:

1. The learning for SOM network is designed to be a Supervised SOM learning algorithm.
2. In this study, Standard Hexagonal lattice structure is being considered as a basis for the enhancement of SOM structure.
3. The proposed algorithms are tested on five standard datasets: Iris, Cancer, XOR, Glass and Pendigits.
4. The performance of the proposed methods are validated based on the Classification Accuracy (CA) and Quantization Error (QE).
5. The programming language to develop the proposed method is in Visual C++ 6.0.

1.7 Significance of the Study

This study investigates the capabilities of enhanced SOM with Particle Swarm Optimization (ESOMPSO) to solve the classification problems. The performance of standard SOM, ESOM, SOMPSO and ESOMPSO using various distance measures: Euclidean distance, Manhattan distance, and Chebyshev distance, are evaluated and compared. The performance of the proposed method is validated and examined to determine its efficiency based on classification accuracy and quantization error. The proposed methods would be alternative solutions for solving complex problems such as pattern classifications, gene expressions, engineering problems and others.

1.8 Organization of the Thesis

This thesis consists of five chapters. Chapter 1 presents the introduction of the study, and Chapter 2 covers the literature review that describes the basic concepts of SOM network, particularly on the architecture and training of Kohonen Self-Organizing Maps (SOM). A related review on past research for classification problems using SOM and PSO algorithm is also presented in this chapter. Chapter 3 provides the framework of the study, and the description of each process. Chapter 4 provides the experimental result of standard SOM, enhanced SOM, SOMPSO and ESOMPSO. The analyses are further validated with various distance measurements and these include Euclidean distance, Manhattan distance and Chebyshev distance on real dataset. Finally, Chapter 5 provides the summary of the research work, research contributions and suggestions for future study.

Furthermore, Sharma and Omlin (2009) utilize a U-matrix of SOM to determine cluster boundaries using PSO algorithm. The authors compared the results with other clustering techniques such as k-means and Hierarchical clustering. However this study doesn't focus on the structure of SOM architecture.

Recently, Ozift *et al.,* (2009) proposed PSO in the optimization of SOM algorithm to reduce the training time without loss of quality in clustering. They stated that the size of lattice is related to the clustering quality of SOM. This optimization technique has successfully reducing the numbers of nodes that finds the BMU for a particular input. By having larger grid size in SOM will invite higher training time. Furthermore, the larger the lattice size, the more nodes should be considered for BMU calculation. This causes higher operating cost for the algorithm.

Due to limitations of the previous studies on focusing the improvement of SOM lattice structure, hence, this study will enhance SOM lattice structure by improving the quality of data classification and labeling with improved hexagonal lattice area proposed by Bariah (2007). Particle Swarm Optimization (PSO) is developed to optimize SOMs' training weights accordingly. The hybridization of SOM-PSO architecture, so-called Enhanced SOM with Particle Swarm Optimization (ESOMPSO) is proposed with improvement on the lattice structure for better classification. The performance of the proposed ESOMPSO is validated based on the classification accuracy and Quantization Errors (QE).

1.3 Problem Statement

This research is focus on enhancement of SOM with PSO for classification problems. The improvement of SOM's lattice area and PSO as the optimization technique can improve the generalization of SOM learning characteristics. Hence, the primary main research question for this study is stated as:

Could the Particle Swarm Optimization enhance the learning capability of Self Organizing Map with Improved Lattice Structure?

This lead to the secondary research questions as below:

i) *How to enhance Self Organizing Map (SOM) with Particle Swarm Optimization (PSO)?*
ii) *How efficient is the improved lattice structure for SOM with different distance measurements?*
iii) *How efficient is the enhanced SOM with PSO for classification problems?*
iv) *How to evaluate the performance of the proposed methods?*

1.4 Research Aim

The aim of this study is to enhance Self-Organizing Map (SOM) learning with improved lattice structure and Particle Swarm Optimization (PSO). This proposed method is evaluated, validated and analyzed in terms of QE and classification accuracy. Different types of distance measurements are implemented on the proposed method to real-valued dataset for classification problems.

1.5 Objectives of the Study

To achieve the aim of the study, few objectives have been identified as below:

1. To develop SOM network with improved hexagonal lattice structure (ESOM).
2. To optimize ESOM with Particle Swarm Optimization (ESOMPSO).
3. To design and evaluate standard SOM, ESOM, standard SOM with PSO (SOMPSO) and ESOMPSO using various distance measurements.
4. To compare the performance of standard SOM, ESOM, SOMPSO and ESOMPSO using standard datasets in terms of quantization errors and accuracy.

1.6 Scope of the Study

Since this study is focusing on the enhancement ESOM with improve lattice structure, the scope are bounded as follow:

1. The learning for SOM network is designed to be a Supervised SOM learning algorithm.
2. In this study, Standard Hexagonal lattice structure is being considered as a basis for the enhancement of SOM structure.
3. The proposed algorithms are tested on five standard datasets: Iris, Cancer, XOR, Glass and Pendigits.
4. The performance of the proposed methods are validated based on the Classification Accuracy (CA) and Quantization Error (QE).
5. The programming language to develop the proposed method is in Visual C++ 6.0.

1.7 Significance of the Study

This study investigates the capabilities of enhanced SOM with Particle Swarm Optimization (ESOMPSO) to solve the classification problems. The performance of standard SOM, ESOM, SOMPSO and ESOMPSO using various distance measures: Euclidean distance, Manhattan distance, and Chebyshev distance, are evaluated and compared. The performance of the proposed method is validated and examined to determine its efficiency based on classification accuracy and quantization error. The proposed methods would be alternative solutions for solving complex problems such as pattern classifications, gene expressions, engineering problems and others.

1.8 Organization of the Thesis

This thesis consists of five chapters. Chapter 1 presents the introduction of the study, and Chapter 2 covers the literature review that describes the basic concepts of SOM network, particularly on the architecture and training of Kohonen Self-Organizing Maps (SOM). A related review on past research for classification problems using SOM and PSO algorithm is also presented in this chapter. Chapter 3 provides the framework of the study, and the description of each process. Chapter 4 provides the experimental result of standard SOM, enhanced SOM, SOMPSO and ESOMPSO. The analyses are further validated with various distance measurements and these include Euclidean distance, Manhattan distance and Chebyshev distance on real dataset. Finally, Chapter 5 provides the summary of the research work, research contributions and suggestions for future study.