AN IMPROVED WAVELET NEURAL NETWORK FOR CLASSIFICATION AND FUNCTION APPROXIMATION

ONG PAULINE

UNIVERSITI SAINS MALAYSIA

2011

AN IMPROVED WAVELET NEURAL NETWORK FOR CLASSIFICATION AND FUNCTION APPROXIMATION

by

ONG PAULINE

Thesis submitted in fulfillment of the requirements
for the Degree of
Doctor of Philosophy

January 2011

ACKNOWLEDGEMENTS

Foremost, I would like to take this opportunity to thank my thesis supervisor, Professor Zarita binti Zainuddin, for introducing to me the theory of wavelets as well as artificial neural networks, and for her expert guidance and encouragement. Her perspective and insight had a profound influence on this thesis.

Special thanks to The Scheme of Graduate Assistant (Teaching) and USM Fellowship under Universiti Sains Malaysia for their generous support which made this research possible.

Last, but certainly not least, I would like to express my gratitude to my parents and family, for their patience, relentless support and sacrifices all these years.

To all others whose names I have failed to mention but who have in one way or another helped in the completion of this thesis, I owe you my humblest gratitude.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

AB	After Breakfast
AD	After Dinner
AL	After Lunch
ALL	Acute Lymphocytic Leukemia
AML	Acute Myelocytic Leukemia
ANNs	Artificial Neural Networks
AT/RT	Atypical Teratoid/Rhabdoid Tumours
BB	Before Breakfast
BD	Before Dinner
BL	Before Lunch
BLY	Burkitt’s Lymphoma
BP	Back-Propagation
BS	Before Sleep
cDNA	Complementary DNA
CG	Classic Glioma
CNS	Central Nervous System Embryonal Tumour
CT	Conditional T-Test Statistic
CWT	Continuous Wavelet Transform
DM	Diabetes Mellitus
DM1	Diabetes Mellitus Type 1
DM2	Diabetes Mellitus Type 2
DSL	Distance Similarity Level
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EEG	Electroencephalogram
EWS	Ewing’s Sarcoma
ERR	Error Criterion Function
FCM	Fuzzy C-Means
FGMM	Fuzzy Gaussian Mixture Model
FWT	Fast Wavelet Transform

FT	Fourier Transform
GAs	Genetic Algorithms
GLO	Glioma
IGT	Impaired Glucose Tolerance
KM	K-Means
LEU	Leukemia
MED	Medulloblastoma
MG	Malignant Glioma
MLPs	Multilayer Perceptrons
MPSDFCM	Modified Point Symmetry-Based Fuzzy C-Means
MPSDKM	Modified Point Symmetry-Based K-Means
MSE	Mean Squared Error
NB	Neuroblastoma
NC	Normal Cerebellum
NG	Non-classic Glioma
NGT	Normal Glucose Tolerance
NNs	Neural Networks
NT	Night
OGTT	Oral Glucose Tolerance Test
OSL	Orientation Similarity Level
PC	Principal Component
PCA	Principal Component Analysis
PNET	Primitive Neuroectodermal
PSD	Point Symmetry Distance
PSDFCM	Point Symmetry-based Fuzzy C-Means
PSDKM	Point Symmetry-based K-Means
RBFNNs	Radial Basis Function Neural Networks
RMS	Rhabdomyosarcoma
RMSE	Root Mean Squared Error
SRBCT	Small Round Blue Cell Tumours
SSL	Symmetry Similarity Level
SVMs	Support Vector Machines
VPC	Bezdek’s Partition Coefficient

wavClust	The Proposed Edge Detection Method
WFT	Windowed Fourier Transform
WHO	World Health Organization
WNNs	Wavelet Neural Networks
WRBF	Weighted Radial Basis Functions
WT	Wavelet Transform

SUATU RANGKAIAN NEURAL WAVELET YANG DITAMBAH BAIK UNTUK PENGELASAN DAN PENGHAMPIRAN FUNGSI

ABSTRAK

Mereka bentuk rangkaian neural wavelet (RNW) dengan sebaiknya adalah penting untuk mencapai prestasi pengitlakan yang optimum. Dalam tesis ini, dua pendekatan yang berbeza dicadangkan untuk meningkatkan kemampuan peramalan RNW. Pertama, jenis fungsi pengaktifan yang digunakan dalam lapisan tersembunyi RNW adalah dipelbagaikan. Kedua, algoritma pengklusteran c-min kabur diperteguh yang dicadangkan—khususnya, algoritma c-min kabur berasaskan simetri titik terubah suai —digunakan dalam memilih lokasi vektor anjakan RNW. Kemudiannya, RNW terubah suai diaplikasikan dalam bidang pengelasan dan penghampiran fungsi. Dalam konteks pengelasan, RNW terubah suai dilaksanakan pada kanser heterogen dan pengelasan diabetes dengan menggunakan lima set data mikrosusunan yang berbeza. Keputusan eksperimen perbandingan menunjukkan bahawa metodologi yang dicadangkan mencapai hampir 100% ketepatan pengelasan dalam ramalan multikelas, yang mendorong pada prestasi superior dengan merujuk kepada algoritma pengklusteran lain. Perbandingan prestasi dengan pengelas lain juga dijalankan. Analisis penilaian menunjukkan bahawa pendekatan yang dicadangkan ini mengatasi prestasi kebanyakan pengelas lain. Dalam konteks penghampiran fungsi, RNW terubah suai diaplikasikan dalam penghampiran lima fungsi yang berbeza. Perbandingan prestasi menunjukkan penambahbaikan yang signifikan dalam ketepatan penghampiran melalui RNW yang dicadangkan. Seterusnya, perbandingan prestasi dengan kaedah lain dalam penghampiran fungsi sesecebis tanda aras yang sama juga dijalankan. Penilaian

menunjukkan kesuperioran pendekatan yang dicadangkan apabila dibandingkan dengan kaedah lain. Suatu kajian RNW yang dicadangkan dalam aplikasi dunia sebenar, iaitu dalam peramalan tahap glukosa darah bagi pesakit diabetes juga dikaji.

Suatu algoritma hibrid baru bagi pengesanan pinggir dibentangkan dalam tesis ini. Algoritma yang terhasil, yang dinamakan, wavClust, kemudiannya diaplikasikan dalam pensegmenan titik imej mikrosusunan. Perbandingan dengan kaedah pensegmenan titik klasik juga dijalankan. Jika dibandingkan dengan kaedah klasik, analisis penilaian menunjukkan bahawa algoritma wavClust yang dicadangkan mampu mensegmen dengan tepat semua bintik berbentuk donut, bintik tak sekata dan bintik dengan keamatan yang pelbagai dan jenis hingar yang berbeza.

AN IMPROVED WAVELET NEURAL NETWORK FOR CLASSIFICATION AND FUNCTION APPROXIMATION

ABSTRACT

Properly designing a wavelet neural network (WNN) is crucial for achieving the optimal generalization performance. In this thesis, two different approaches were proposed for improving the predictive capability of WNNs. First, the types of activation functions used in the hidden layer of the WNN were varied. Second, the proposed enhanced fuzzy c-means clustering algorithm—specifically, the modified point symmetry-based fuzzy c-means (MPSDFCM) algorithm—was employed in selecting the locations of the translation vectors of the WNN. The modified WNN was then applied in the areas of classification and function approximation. In the context of classification, the modified WNN was implemented to heterogeneous cancer and diabetes classification using five different microarray datasets. The comparative experimental results showed that the proposed methodology achieved an almost 100% classification accuracy in multiclass prediction, leading to superior performance with respect to other clustering algorithms. Performance comparisons with other classifiers were made. An assessment analysis showed that this proposed approach outperformed most of the other classifiers. In the context of function approximation, the modified WNN was applied in approximating five different functions. Performance comparisons indicated significant improvement in the approximation accuracy was accomplished by the proposed WNN. Subsequently, performance comparisons with other methods in approximating the same benchmark piecewise function were made. Evaluation demonstrated the superiority of the proposed approach when compared with other

methods. A study of the proposed WNN in a real-world application, i.e. prediction of blood glucose level for diabetics was also investigated.

A novel hybrid algorithm for edge detection was presented in this thesis. The resulting algorithm, namely, wavClust, was then applied in the microarray image spot segmentation. Comparisons with the classical spot segmentation methods were made. Assessment analysis showed that the proposed wavClust algorithm was able to segment all the donut-shaped spot, irregular spot and spots with intensity variations and different noise types accurately, compared to the classical methods.

CHAPTER 1

INTRODUCTION

1.1 PRELIMINARIES

The invention of the abacus 2000 years ago is the starting point of the evolution of the computer. Simply using a wooden rack holding two horizontal strings with beads strung on them, basic arithmetic problems can be solved by moving the beads around, following the simple programming rules memorized by the user. A few centuries later, through further innovations and improvement, we have the invention of another wonderful machine: the computer. Life has never been the same since.

The advent of the computer is a blessing to the emergence of the artificial neural networks (ANNs). The study of the human brain dates back several centuries, inspired by the ways of how humans and living creatures struggle to survive in a challenging environment. The human brain possesses incredible characteristics. It is robust and fault-tolerant; able to adaptively adjust to new environments through learning; capable in dealing with diffused information; and able to process multiple sets of parallel information at the same time. Indeed, the human brain is amazing. It can recognize different faces, discern noise from music and differentiate various types of shapes. Thus, it is not surprising that the members of the scientific community endeavor to unlock its mystery. Although researches on the human brain have been ongoing for a long time, scientists have only recently been able to successfully simulate the thinking process of the human brain after the invention of the computer.

The aspiration of scientists to emulate the complex functions of the human brain has resulted in a new challenging research area, namely Artificial Intelligence, where ANNs are merely one of the facets of this interesting field. Attempting to preserve the desirable features of the human brain, mimicking its structures and understanding how it operates constitute the core development of ANNs.

An extensive understanding of the constituents and the manners in which the human brain processes information is crucial and necessary before constructing the building blocks of ANNs. Yet, the struggle to understand the exact workings of the human brain is not an easy task. The human brain is the most complex organism, made up of a biological nervous system of organized assembly of cells interconnected through synapses. Fortunately, in 1911, Ramón y Cajăl came up with the idea of introducing "neurons" as the elementary structural components of the brain. His contribution has helped scientists better understand the secrets of the human brain (López et al., 2006).

It is believed that the human brain is composed of approximately 10 billion neurons. Each neuron can connect directly with up to 200,000 other neurons (though 1,000 to 10,000 is typical) (Gopal, 2009). The ability to think intellectually, to remember, to learn, and to experience sensations is attributed to the enormous number of neurons and the vast interconnections between them.

Basically, a biological neuron consists of four main components, namely, dendrites, soma, axon and the synapse. Soma forms the main body of a neuron. The dendrites of a neuron are hair-like filaments, which branch out from the soma. An axon is a long, slender projection that connects the soma with the terminal branches of the axon. There are a number of synapses at the end of the terminal branches, which usually connect to the dendrites of the other neurons. The structure of a typical biological neuron is illustrated in Figure 1.1.

The nerve impulse transmission of a neuron starts at the dendrites. Dendrites, the receptive zone, collect signals from the other neurons. The signals are converted into electrical impulses and propagated away from the soma via the axon, the transmission line, until it reaches the synapses. At the synapses, communication with the other neurons occur (Figure 1.2).

At the synapses, the nerve impulse will be converted into electrical effects that cause either excitation or inhibition in the neuron which is connected to it. When a neuron sends an excitatory signal to the neuron it is connected to, this signal will be summed up with all the other inputs of that neuron. If it exceeds a given threshold, the target neuron will fire an action potential. If it is lower than the threshold, no action potential takes place.

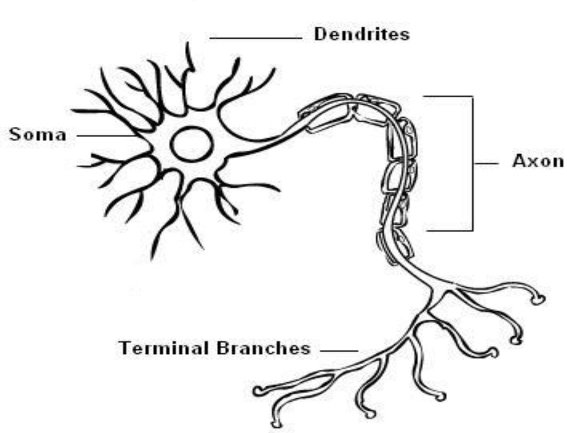


Figure 1.1: Schematic diagram of a neuron

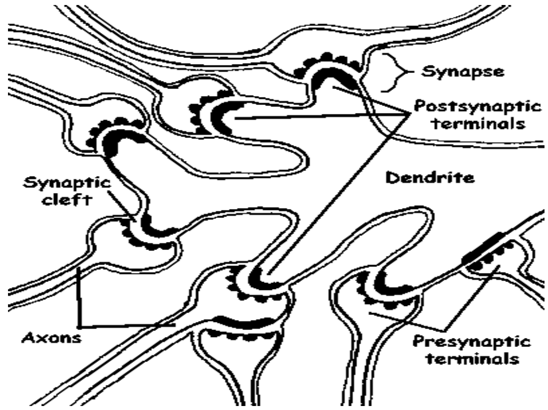


Figure 1.2: Connection between the synapses and the dendrites of other neurons

In short, a biological neuron receives signals from other cells, perceives it and makes appropriate decisions. It combines the received signals in some way, performs a generally non-linear operation on the result of combination, and then outputs the final result to the other neurons.

An artificial neuron, which is the core processing element of an ANN, is a mathematical function that imitates how the biological neurons operate. The first artificial neuron model, developed by McCulloch and Pitts (1943) in 1943, is shown in Figure 1.3.

The flows of the neuron in the McColluch and Pitts model are similar to the processes involved in the biological neurons. The inputs received by the artificial neuron from the other neurons (representing the nerve impulses received by the dendrites) are weighted in order to modulate the strength of the input. The sum of the weighted inputs is then fed through a transfer function to generate an output signal, which is then fed into the other neurons. The transfer functions could be in the form of sigmoid functions, piecewise linear functions or step functions.

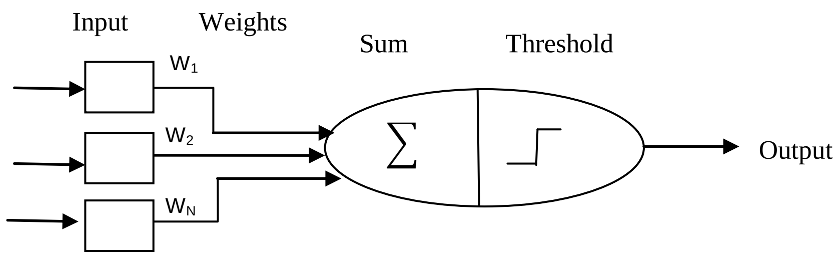


Figure 1.3: McColluch and Pitts model

Following McColluch and Pitts's pioneering work, a series of great advances has arisen in the improvement of ANNs over the past few decades. The timeline for the brief history of the major development in ANNs is summarized in Table 1.1 (Haykin, 1999; Iyengar, 2002).

Table 1.1: Development of ANNs over the past few decades

Year	**Researchers**	**Findings**
1943	McColluch and Pitts	The first artificial neuron was developed.
1949	Donald Hebb	When a human learns a different task, the connectivity of the brain is changed continually. Based on the ways of how a human learns, the learning rule of the synaptic modification was proposed. It describes how the neural pathways are strengthened each time they are used, which is a concept identical to the learning process in the human brain.
1954	Marvin Minsky	A classical paper entitled "Steps Toward Artificial Intelligence" was published, where a large section in this paper is now termed as neural network.
1958	Frank Rosenblatt	A novel supervised learning method, named as perceptron convergence theorem was introduced.

Table 1.1: Continued

Year	Researchers	Findings
1960	Widrow and Hoff	The least mean-square algorithm was formulated in the models of ADALINE and MADELINE.
1969	Minsky and Papert	The declaration of the limitation of single layer perceptrons was made, where it led to the inactive research period of ANNs.
1982	Hopfield	The Hopfield model, which is a recurrent network with symmetric synaptic connections, was introduced.
1982	Kohonen	A novel structure of ANNs, namely, self-organizing maps which uses a one or two dimensional lattice structure was proposed.
1986	Rumelhart, Hinton and Williams	The development of the back-propagation learning algorithm was reported.
1988	Broomhead and Lowe	A single hidden layer of radial basis function neural network was developed.
1992	Vapnik	The first support vector machine was invented.
1992	Zhang and Benveniste	Wavelet neural networks (WNNs) are reported.
Today		With the dawn of the advances in the technology today, the researches in ANNs are studied widely all around the world.

The following section provides an introduction to ANNs, and how the problems encountered in the typical ANNs inspire the formulation of WNNs. Next, the motivations for developing an improved WNN in this research are presented. This is followed by the research objectives, research scope and lastly, an overview of the organization of this thesis is given.

1.2 INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

An ANN, or normally just referred as neural networks (NNs), is made up of interconnecting artificial neurons, which uses mathematical or computational models for information processing. A schematic diagram of an ANN with two hidden layers is given in Figure 1.4.

The basic component of an ANN is the artificial neuron. The combination of the artificial neurons forms a layer, namely, the input layer, the hidden layer and the output layer. As shown in Figure 1.4, every artificial neuron is connected to the succeeding layer, where there is a synaptic weight associated with each neuron.

Two important attention-grabbing characteristics of ANNs are (Haykin, 1999):

- It possesses self-learning ability, and
- It uses simple computational operations in reaching the solution for a highly complex, mathematically ill-defined and non-linear problems

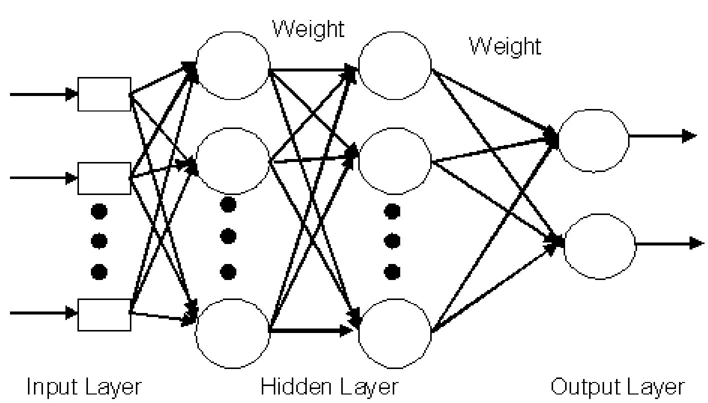


Figure 1.4: Schematic diagram of an ANN

Distinct from the computer systems which are only capable in doing the precised programmed instructions, the ANNs are able to classify the objects that it has never seen before into classes or even predict the future based on past experience. These fascinating characteristics are attributed to the learning ability of ANNs. Much like how humans learn from experience, ANNs also learn from examples. The learning of ANNs involves the adjustment of the synaptic weight connections that exist between neurons. It changes its network structure adaptively based on external and internal information that flows through the network during the learning phase.

For example, in the supervised learning method, a collection of samples are fed into the ANNs with predefined target output. After the learning phase, the output achieved from the ANNs is compared with the predefined target output. If the undesired output is obtained, the altering to the synaptic connection weight is made, such that the error within the network results and the target output is minimized.

It is not surprising that the ANNs with the property of massive parallel distributed structure and the ability to generalize reasonable output for the unseen data during the learning phase make it a suitable tool for solving large-scale problems where the physical processes are highly complex and ill-defined. Furthermore, there are difficulties in solving the problems that do not have an algorithm solution or the solution is too complex to obtain by using the conventional methods. This is where the ANNs play its role.

ANNs outshine other conventional approaches due to its adaptive learning capability. During the learning phase, an ANN learns the underlying relationship between the input and output data. It has been found that, an ANN provides correct solutions even though the input data is contaminated by noise, when properly trained (Juang et al., 2009). ANNs perform excellently in pattern recognition and classification problems where correct decisions are being made when the imprecise input data are present.

Due to its fascinating features of robustness, fault tolerance, adaptively learning and highly parallel capabilities, ANNs have been implemented extensively in our daily life (Adeli and Panakkat, 2009; Adjeroh et al., 2006; Balasubramanian et al., 2009; Crone and Kourentzes, 2010; Du, 2010; Ebrahimzadeh and Ranaee, 2010; Emili et al., 2008; Huang and Wu, 2008; Jiao, 2010; Juan and Julian, 2006; Kim and Cho, 2006; Lee, 2008; Lee and Ko, 2009; Li et al., 2009; Lin and Hsieh, 2009; Lu, 2010; Melin et al., 2008; Pajares et al., 2010; Suk et al., 2010; Sun and Tien, 2008; Übeyli, 2008; Übeyli, 2009; Wang et al., 2009).

The applications of the ANNs fall within the following broad categories:

a. Function approximation and regression analysis, including the time series prediction and modeling
b. Classification, including pattern recognition and decision making
c. Data processing, including data compression, filtering and clustering

The implementations of ANNs in the real-world problems can be discriminated further into more detailed areas of application. Some examples of the specific paradigm are given for each area:

a. Finance and Commerce
    - Forecasting of stock price index
    - Forecasting the exchange rate
b. Medical
    - Classification of heterogeneous cancer subtypes
    - Denoising and compression of medical image
    - Classification of EEG signals
c. Environment
    - Prediction of the pollutant concentrations
    - Forecasting of the storm surge at the costal line
d. Pattern Recognition
    - Face recognition
    - Handwriting recognition
    - Fingerprint recognition
e. Engineering
    - Fault detection and diagnosis
    - Control system

1.3 PROBLEM STATEMENT

Accuracy is seen as important by everyone, in every aspect, in view of the fact that it is always correlated with the decision-making. For example, when the heterogeneity of cancer could not be differentiated through its morphological appearance, classification of the heterogeneous cancer accurately is tremendously crucial, since a correct classification enables the maximizing of efficacy and minimizing of toxicity in the therapy. Since numerous application problems are highly dependent on the accuracy of the ANNs, improving the prediction and approximation accuracy of the ANNs is the main concern of this thesis.

To date, multilayer perceptrons (MLPs) (Ghate and Dudul, 2010; Haykin, 1999; Hornik et al., 1989; Zainuddin and Ong, 2007), support vector machines (SVMs) (Huang and Wu, 2008; Peng et al., 2003; Polat et al., 2008; Shim et al, 2009; Xian and Zeng, 2009; Xu et al., 2009; Zhang et al., 2006), radial basis function neural networks (RBFNNs) (Balasubramanian et al., 2009; Broomhead and Lowe, 1900; Lee and Ko, 2009; Moody and Darken, 1989) and fuzzy neural networks (Dazzi et al., 2001; Juang et al., 2009; Othman and Yao, 2007; Polat and Günes, 2007;) are some of the popular ANNs that have been implemented successfully in a vast variety of applications. Even though MLPs along with the backpropagation learning algorithm are the most popular type of ANNs for practical situations, the deficiencies of a MLP's multilayer structure and its use of a global activation function and a slow learning algorithm have limited its use in practice. Specifically, MLPs have difficulties reaching the global minimum in a complex search space, are time-consuming and fail to converge when high nonlinearities exist; these issues have deteriorated their generalization capability to achieve superior accuracy (Lin and Tsai, 2008; Yao et al., 1996; Zainuddin and Ong, 2010). Apart from that, there are issues that remain unresolved in selecting the appropriate centers for RBFNNs (Liao, 2010; Staiano et al., 2006; Zainuddin and Lye, 2010), membership function for fuzzy neural networks (Hsu et al., 2008) and kernels for SVMs (Chen et al., 2007; Ju at al., 2009; Kazuhiri, 2008; Wu, 2010).

WNNs, as one of the facets of ANNs research filed, had been introduced by Zhang and Benveniste as a vital alternative to MLPs that overcome their shortcomings (Zhang, 1997; Zhang and Benveniste, 1992) where it differs from other ANNs with the constitution of the wavelet activation function in the hidden nodes. They had proven that the proposed WNNs possess the eye-catching uniqueness of (Zhang and Benveniste, 1992):

- It preserves the universal approximator property
- Explicit link exists between the network coefficients and wavelet transform
- It achieves the same quality of approximation with a network of reduced size

Due to the advantages of WNNs as universal approximators, the fact that they have more compact topology than other neural networks and their fast learning speed owing to the constitution of the localized wavelet activation function in the hidden layer, WNNs had received much attention from other researchers and have been used extensively to solve numerous real world problems such as face recognition, time-series prediction, pattern classification and system identification (Alan et al., 2006; Avci, 2007; Becerra et al., 2005; Biswal et al., 2009; Cao et al., 2010; Chaohan et al., 2009; Cui et al., 2005; Gutés et al., 2006; Kumar et al., 2008; Lin, 2009; Pan et al., 2008; Subasi, 2005; Zainuddin and Ong, 2010).

Various issues have been addressed in the WNNs studies, which include adjusting the connection weights by employing different learning algorithms for accelerating the convergence of WNNs (Chen et al., 2006; Lin and Tsai, 2008; Yao et al., 1996; Zhang, 2007), making alterations in the network architecture (Wan et al., 2004), introducing variation in the types of activation functions used in the hidden layer (Mohd Idris et al., 2009; Wajdi et al., 2005, Zainuddin and Ong, 2007), and modifying the wavelet function parameters, a process that involves proper initialization of the translation and dilation vectors (Lin, 2009; Oussar and Dreyfus, 2000; Zhang et al., 2006). In this thesis, optimizing the convergence characteristic and improving the generalization ability of WNNs by emphasizing the choice of a proper wavelet family as the activation functions in the WNNs hidden layer and proper initialization of translation vectors are the main concern.

There is heightened understanding that selecting the appropriate activation function in hidden layer is as crucial as the neural network architecture and learning algorithm (Dutch and Jankowski , 1999; Gougam et al., 2008). Being the first proposed activation function, it had been proven in the universal approximation theorem that a single hidden layer feedforward neural network with sigmoid activation functions can approximate any continuous, multivariate function to any desired degree of accuracy with a finite number of neurons (Cybenko, 1989). However, the main limitation of sigmoid functions is that they span over the whole input space. The alteration of the weight vectors and other parameters involves all the activation functions, and thus its training is time-consuming, and unavoidably, it achieves much more exploration errors. As opposed to the popular sigmoid activation functions, wavelets in hidden layer respond only to a local region of the space of input values. Due to its fast-decaying characteristic in a short finite length interval, fewer non-negligible coefficients are generated. Thus, modification on the weight vectors involves only a small number of parameters and as a result it leads to fast convergence characteristic and generalization capability (Oysal et al., 2005).

Systematic investigation on the suitability of employing various wavelets as the activation functions (for example, B-spline wavelet, Gabor wavelet, Gaussian wavelet, Mexican Hat, Morlet and Shannon wavelet (Banakar and Azeem, 2008; Chaohan et al., 2009; Cui et al., 2005; Esen at al., 2009; Krüger and Sommer, 2002; Lin et al., 2008; Wang et al., 2005;)) for WNNs in the literature showed its feasibility and validity. However, there is no a priori explanation on why they should be the most favorable in all situations. Simulation results are highly relied on the network architecture, learning algorithm, parameter initialization, and also on the dataset used. Unfortunately usually the essential attributes of the dataset is unknown and vary for different problems, a wavelet activation function that is well-suited for all the cases does not exist. The choice of the wavelet families is problem-dependent (Mohd Idris et al., 2009; Zainuddin and Ong, 2007, Zainuddin and Ong, 2010). Since there is no established rule of thumb in determining which particular wavelet to be employed as the activation function, this research hopes to make some suggestions.

WNNs update their connection weights and parameters iteratively through learning. During the learning process of a WNN, the selection of the numbers and the locations of the translation vectors are particularly crucial. An appropriate initialization of the translation parameter will do a good job of reflecting the essential attributes of the input samples, which is important for finding an optimal solution. Increasing the number of hidden nodes leads to over-fitting and computational complexity. Thus, assigning an appropriate number of hidden nodes simplifies the process.

Several approaches have been suggested in choosing the appropriate WNNs translation vector, including using an explicit expression (Cao and Lin, 2008; Cao et al., 2010; Cui et al., 2005; Gutés et al., 2006; Gutiérrez et al., 2008; Lin, 2009; Oussar and Dreyfus, 2000; Srivastava et al., 2005; Zhang et al., 2004), hierarchical clustering (Wei et al., 2004), SVMs (Zhang et al., 2006), genetic algorithms (GAs) (Kim et al., 2002) and k-means clustering (Hwang et al., 2000). Although such improvements are certainly noteworthy, initializing the translation vector is always an open question. Therefore, a novel clustering algorithm—the modified point symmetry-based fuzzy c-means (MPSDFCM) was proposed in this study—as an alternative to the existing WNNs translation parameter initialization approaches.

1.4 OBJECTIVES OF THESIS

The main thrust of this research is geared towards an improved WNN model, incorporating the MPSDFCM algorithm and different wavelet activation functions, for prediction purposes as well as pattern classification.

The objectives of this thesis are as follows:

- To implement different types of activation functions in WNNs for the purpose of improving its generalization ability
- To develop and apply a novel method in the parameter initialization of WNNs for the purpose of optimizing its generalization ability
- To assess the potential benefits of the proposed enhanced WNNs in heterogeneous cancer detection and blood glucose concentration prediction

- To develop and implement a new scheme in edge detection based on the hybridization of wavelet transform and clustering algorithm

1.5 SCOPE AND ORGANIZATION OF THESIS

The scope of this research studies the development of an improved WNN model. Particular focus is placed on the translation parameter initialization using MPSDFCM algorithm, and the choice of wavelet activation function in the hidden nodes. The effectiveness of the proposed WNNs with MPSDFCM initialization method in the context of function approximation and classification is examined through empirical approaches, with simulated as well as real-world dataset. The applicability of the proposed WNNs is extended to practical applications in bioinformatics, which include the real-world medical diagnosis, in the domain of heterogeneous cancer classification and the prediction of blood glucose concentration for a diabetic.

The thesis will begin with Chapter 1 with an introduction to the evolvement of ANNs and its development history. The characteristics of the ANNs are explored and examples on its implementation in real-world applications are given. The objectives, scope and organization of this thesis are then described.

Chapter 2 will provide a review of WNNs, which begins with an introduction to wavelet and wavelet transform. The preface for architecture, learning algorithm and parameter initialization of WNNs are given. Its competence is briefly touched on. The characteristics of the WNNs are reviewed and the need to enhance the WNNs from different aspects is then explored. Implementation of WNNs in the real-world applications is then given.

A review of the cluster analysis is given in Chapter 3. It begins with the introduction to the classical clustering algorithms, and this is followed by the exploration of their pro and cons. The concept of point symmetry distance (PSD) is introduced, and its improvement over the conventional clustering algorithm is then explored. The merit of a novel operator, i.e. symmetry similarity level (SSL) operator is then described. Motivated by the SSL operator, a novel clustering algorithm-specifically, the modified point symmetry-based fuzzy c-means (MPSDFCM) algorithm is proposed. The description and the implementation of the proposed MPSDFCM, which is the core of this thesis is then given.

Chapter 4 will begin with an overview of the microarray experiment. The research background is then provided, and it will lead to the need of developing an enhanced classifier to address the problems in multiclass cancer classification. Its implementation in the heterogeneous cancer classification using benchmark microarray dataset was studied. Assessment analysis of the proposed enhanced classifier with other well-developed classifiers is then presented. It will be shown that the modification on the WNNs improves its prediction capability effectively and significantly.

The enhancement of the WNNs approximation capability sets the tone for Chapter 5. The proposed WNNs are applied in function approximation problems, where different types of functions are studied. Evaluations by varying types of activation function and initialization approach in terms of mean squared error are made. A real world application, i.e. prediction of blood glucose level for diabetics is then explored. The capability of the proposed WNNs in approximating the blood concentration at the end of interval is assessed further, by comparing its prediction accuracy with other popular ANNs.

Edge detection is one of the main concerns in image analysis. The image quality issue in the microarray experiment needs more improvement. Chapter 6 will begin with an introduction for microarray image analysis. It is followed by a brief review on the edge detection based on the wavelet approach. A new scheme of edge detection with hybridization on the wavelet and clustering algorithm is then proposed. Its implementation in microarray spot segmentation is studied. Comparison of the segmentation results with other existing segmentation method will round up this chapter.

Finally, refinements on WNNs and edge detection are proposed for future research direction. A brief elucidation on recommendations for further research and conclusions of this research work will conclude the thesis. Further details and images are provided as Appendices.

CHAPTER 2

WAVELET AND WAVELET NEURAL NETWORKS

2.1 INTRODUCTION

Wavelet analysis has been studied rigorously in a number of disciplines in mathematics, quantum physics and electrical engineering over the past few decades. These active researches have led to the emergence of wavelet as one of the fastest growing fields with vast new applications, ranging from image compression, voice recognition to earthquake prediction. Owing to its fascinating characteristics such as fast-decaying, compact support, smooth and regular, wavelet has been substituted as the activation function in the WNNs, which greatly overcomes the shortcomings in MLPs.

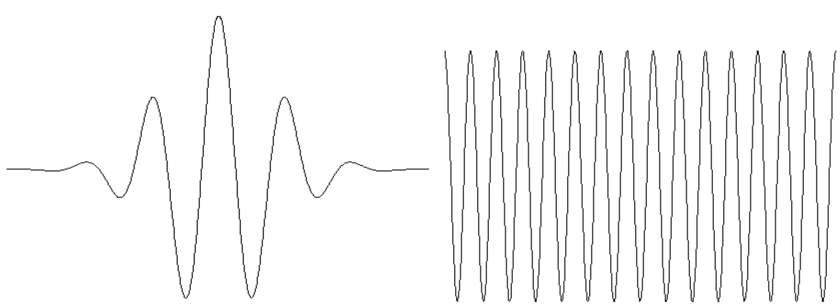
In this chapter, the fundamental aspects of wavelet analysis and WNNs will be reviewed in a number of sections. Specifically, this chapter begins with an overview of the wavelet. The ideas behind discrete wavelet transform (DWT), with the discussion on its improvements over the Fourier transform (FT), windowed Fourier transform (WFT) and continuous wavelet transform (CWT) is presented. Next, we proceed to the brief review of the DWT in a vast variety of practical applications. In the second section, an overview of the theory of WNNs, which includes the architecture design, learning process and parameter initialization of a WNN, is given. Finally, a literature study on the implementation of WNNs in numerous real-world problems is presented at the end of this chapter.

2.2 INTRODUCTION TO WAVELET

The word wavelet, originates from French word *ondelette* which means “small wave”, was introduced by Morlet and Grossman in 1980. The term “small” here refers to the condition that the wavelet function is of finite length, which means compactly supported, whereas the term “wave” implies that the function is oscillatory.

A wavelet is a kind of mathematical function that satisfies certain requirements, such as integrates to zero, oscillates and well localized in time. It is used to represent other functions by cutting it into different frequency components, and then study each component with a resolution that matches its scale. Thus, it has the ability to allow the time and frequency analysis simultaneously, which makes it a suitable tool for studying the transient, non-stationary and time-varying signals.

A wavelet is not identical to a wave. A wave is an oscillating function of time and it is periodic. In contrast, wavelets are localized waves. An example of a wavelet and a wave (sinusoid) is shown in Figure 2.1.



(a) (b)

Figure 2.1: A schematic diagram of (a) a wavelet; and (b) a wave (sinusoid)

It can be observed that a wavelet has the oscillating wave-like characteristic like a wave (sinusoid). Waves are smooth and regular, but wavelets are irregular and might be asymmetric in shape. Unlike sinusoid which oscillates with equal amplitude from minus to plus infinity, wavelet has a finite length and it decays quickly toward zero when their limits approach minus to plus infinity. Therefore, wavelets have finite energy which concentrates around a point.

Wavelets are generated from a finite-length or fast-decaying oscillating mother wavelet $\psi(t)$, with an average value of zero

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{2.1}$$

The family of functions can be obtained by shifting and scaling of this mother wavelet as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), a, b \in R, a \neq 0 \tag{2.2}$$

where a and b are scaling and translation parameters, $\frac{1}{\sqrt{a}}$ is energy normalization so that the transformed signal will have the same energy across the scales. The mother wavelet gets its term "mother", since the functions with different support regions in the transformation process are generated from it. Thus, mother wavelet is the prototype for deriving other window functions, with different values of scaling and translation parameters.

A wavelet function $\psi(t)$ must satisfy the admissibility condition in order for it to be the mother wavelet. A wavelet function $\psi(t)$ is said to be admissible if its Fourier Transform, namely

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(u) e^{-i2\pi f u} du \tag{2.3}$$

satisfies the admissibility condition

$$0 < C_\psi = \int_0^\infty \frac{|\Psi(f)|^2}{f} df < \infty \tag{2.4}$$

2.3 PROPERTIES OF WAVELETS

There are some important properties of wavelets that make it a useful tool in real world applications.

Vanishing Moment: A wavelet function $\psi(t)$ is said to have *p* vanishing moments, if

$$\int_{-\infty}^{+\infty} x^k \psi(t) dt = 0, 0 \leq \mathrm{k} < p \tag{2.5}$$

If a wavelet function $\psi(t)$ has larger vanishing moments, the wavelet coefficients of a function $f(t)$ are much smaller on a larger scale *j*, where this wavelet is said will have more compressive power.

Regularity: The regularity of a wavelet function $\psi(t)$ is related to the vanishing moments. To have regularity more than *n*, $\psi(t)$ must has at least *n+1* vanishing moments. Hence, the more regularity of a wavelet, the more vanishing moments it has, and the smoother it is.

Smoothness: The smoothness of a wavelet function $\psi(t)$ is determined by the number of vanishing moments, where the smoother a wavelet function is, it has a better reconstruction property. A smoother wavelet function $\psi(t)$ will have fewer non-negligible wavelet coefficients, which means that it produces a large number of wavelet coefficients that are close to zero, which is essential for noise removal and image compression.

Compact Support: The compact support of a wavelet function $\psi(t)$ is the maximal interval of which the wavelet function vanishes outside of this finite interval. If the size of the compact support is smaller, there are fewer of the high amplitude wavelet coefficients.

Symmetry: Symmetry is also called as linear phase in the language if engineering, which is an important characteristic in image processing, where a wavelet which is symmetric in shape produces less visual artifacts than an asymmetric wavelet. The absence of this property can lead to a phase distortion.

2.4 A BRIEF HISTORICAL PERSPECTIVE OF WAVELETS

- ***Pre-1930***

  Notable contribution to the development of wavelet analysis first originated from Alfred Haar's work in the early of 20th century, where the Haar wavelet function was first mentioned. Haar wavelet has compact support; however it is not continuously differentiable, where this shortcoming somewhat limits its applications (Burrus, 1998; Eugenio, 1996; Holschneider, 1995).

- ***The 1930s***

  After Haar's contribution to wavelets, Paul Levy (1930) discovered that a Haar basis function is a better tool than Fourier basis functions when dealing with the small details in Brownian motion.

- ***Post 1980***

  The next major advancement of wavelets comes from Jean Morlet and Alex Grossman in the year 1981. They found that a signal can be transformed into wavelet forms and then be reverted back into its original signal without loss of information. This property contributes greatly in data compression, where making a small change in the wavelets will cause a small change in the original signal.

  Stephane Mallat (Mallat, 1989) continued the work on wavelet analysis by discovering the relationships between the quadrature mirror filters and pyramid algorithms, laying the foundation to the fast wavelet transform (FWT).

  The idea of multiresolution analysis of wavelets which was a big jump in the research of wavelet analysis comes from Yves Meyer and Stephane Mallat (Mallat, 1989). The idea of looking at a signal at different scales of resolution is beneficial to the next important discovery by Ingrid Daubechies. The final great salvo in wavelet analysis happened at around 1988, where Ingrid Daubechies created a new family of wavelets, called Daubechies wavelets, by using the idea of multiresolution analysis. Daubechies wavelet families are both continuously differentiable and have compact support. These properties have made it the keystone in wavelet applications today.

The development of wavelet analysis is highly related to the Fourier transform (FT). Before the emergence of wavelet analysis, there are vast applications of FT in signal analysis. However, there are certain shortcomings in FT in addressing the non-stationary signals. To overcome these drawbacks in FT, development of wavelet analysis which takes hints from Fourier analysis appeared as a better resolution than FT in signal processing.

2.5 FROM FOURIER TO WAVELET ANALYSIS

Signal analysis has benefited from mathematical transformation, such as FT and the more recently wavelet transform (WT). Mathematical transformations are always applied to raw signals in order to obtain the hidden information that is unavailable in raw signals.

2.5.1 FOURIER TRANSFORM

It is best to describe the wavelets by starting with FT. Among all the mathematical transformation methods, FT is the most popular. Since the early of 1800, Joseph Fourier had discovered that any periodic functions can be expressed as the superposition of sine and cosine functions. Any 2π periodic function $f(x)$ is the sum of

$$a_0 + \sum_{k=1}^{\infty}(a_k \cos(kx) + b_k \sin(kx)) \tag{2.6}$$

of its Fourier series, where

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)dx,\ a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(kx)dx,\ b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(kx)dx \tag{2.7}$$

Since most of the real-world functions are not in the type of periodic functions, such as the sound of a motor that speeds up, the spoken word and the melody of a song, many years after this remarkable property of periodic functions were discovered, Fourier’s idea was extended to analyze the non-periodic functions that have great changes over the time.

However, FT struggles to reproduce the transient signals or signals with abrupt changes due to its frequency-amplitude representation of the signal, which is