CHAPTER I

INTRODUCTION

An artificial neural network (ANN) is a mathematical model or computational model. This model is based on biological neural networks. Moreover, it consists of an interconnected group of artificial neurons and processes data using a connection approach to computation. Normally, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning stage. The ANN can be used to be a model complex relationships between inputs and outputs. Nowadays, the ANN has been used as an alternative tool applying in many tasks. This technique is for solving various problems in different fields of science and engineer such as function approximation, classification, pattern recognition, diagnosis, etc.

Neural network learning is classified into three types; supervised learning, unsupervised learning, and reinforcement learning. This research is focusing on supervised learning. The Supervised learning is a machine learning technique for creating a function from training data. Pairs of input objects and desired outputs are the major factors of the training data. The output of the function can be a continuous value (Approximation), or can predict a class label of the input object (Classification). After having seen a number of training examples, the supervised learner has been designed to predict the value of the function for any valid input object. To achieve this, the learning process has to generalize from the presented data to unseen situations.

Three things to consider in the learning process are network architecture, learning model, and training algorithm. The proper training algorithm will make the network efficiently from the cost function. The multilayer perceptron is widely used as a network architecture, and commonly backpropagation (BP) is used as a learning algorithm. The BP also uses the mean squared error (MSE) as a cost function. The cost function in the BP algorithm adjust weights until the output satisfy the target with the acceptable error criterion. However, the output and the target values may not have parallel differences as some output values can be higher or lower than the target. Furthermore, the output values may fluctuate within an acceptable error criterion. So the BP approach results in an ineffective forecasting, and causes the slow convergent rate or nonconvergence owing to the excess of the acceptable error.

1.1 Statement of Problems

Two subproblems are

1. Is there any cost function such that, the trained network can reduce the training time and still learning the main behavior of the target? According to the principles of backpropagation learning algorithm, the problem of this method was found that the output can be higher or lower than the target value within an acceptable error criterion when using the mean square error cost function. Therefore, a new cost function that enables forecasting similar behavior should be proposed. How should a new cost function be in order to capture the behavior of the movement, regardless the actual output value?
2. Is there any application that applies a neural cellular automata to learn the natural phenomenon with minimum training time?

We have implemented the back-propagation neural network (BPNN) with a structure of feedforward multilayer perceptron for simulating the diffusion characteristics of water. The cost function used for backpropagation learning algorithm is applied for minimizing the difference between target and output values obtained from BPNN of every training pattern by using gradient descent method for adjusting the weight. The BPNN becomes stable when the selected error criterion defined by cost function is reached. However, the output and target values may not have a parallel differences as some output values can be higher or lower than the target value, which usually causes the slow convergence rate of BPNN. Therefore, the first subproblem of this work is aimed to propose the a cost function capable for forecasting similar behavior.

In the first sub-problem, proposing a new cost function, J. G. Juang[1] investigated effects of using different neural network structures and cost functions in locomotion control by using mean square error. Normally, position and velocity are considered, but this research concern only the effects of position parameters in the cost function. Lazaro et al. [2] studied a new cost function for binary classification problems based on the distributions of the estimated output for each class. The study used the inverse of cost function to reduce the overlap between distributions, instead of mean square error. The inverse of the cost function can be regarded as a divergence measurement between estimated distributions.

To avoid the problem of the variation of the output value which can be higher or lower than target, S. F. Crone[3, 4] proposed a set of asymmetric cost functions as objective functions for neural network training and obtaining superior forecasts even for white noise time series.

S. Weaver[5] proposed the localizing learning to improve supervised learning algorithms by reducing of a objective cost function that combines the approximation error and a term that measures interference. Supervised learning of smoothing parameters in image restoration by regularization under cellular neural networks framework mapped the cost function of the optimization problem into the Lyapunov function [6]. Other studies [7, 8, 9, 10, 11, 12, 13, 14, 15] adjusted weight algorithm and cost function to speed up the learning rate or to be more efficient.

From the reviewed papers and the dot product concept, this research proposes a new cost function which can imitate target and output behavior such that they are parallel.

For the second sub-problem, Baensch, Morin and Nochetto[16]were able to calculate the time evolution of some simple geometrical shapes in two and three dimensions and present the computational results by using animation[17]. They applied a finite element method for surface diffusion by using the partial differential equations (PDE) which was sometimes too difficult to find the initial value and the boundary condition, and it was also complicated to select the appropriate initial values that give accurate results. Further, the accuracy of the prediction highly depends on the boundary conditions, and other existing parameters in the system. Recently, a new methodology called parallel distributed model[19] has been proposed to model a variety of problems. Typical examples are shown in the reports [18, 19] on the dynamical prediction of a system such as flood wave propagation. Their studies concerned only one output variable which can be either flood elevation or discharge rate as a function of time. Under this limitation, visualization and animation of a dynamical behavior were not possible. The attempt to simulate physics-based graphic model by using neural network was reported in [20]. This study emphasized the prediction of physical values to be used for animating the visualization afterwards. From the above reviewed papers and the neural cellular automata principle, this research combines all these ideas and then applies to simulate the color diffusion.

The rest of the dissertation is organized as follows. Section 1.2 reviews the related literatures. Chapter II summarizes the theoretical background involving the problem of supervised learning neural network. Chapter III explains the proposed method, model and the learning algorithm of the proposed cost function. Chapter IV discusses and compares all the experimental results. Chapter V concludes the dissertation.

1.2 Literature Review of Literature Related to Cost Functions

J.G. Juang[1] investigated the effects of using different neural network structures and cost functions in locomotion control. Simulations focused on refinement and a thorough understanding of an artificial intelligent learning scheme. This scheme used a neural network controller with backpropagation through time learning rule. Through learning, the controller could generate locomotion trajectory along a pre-defined path. Different issues regarding the scheme had been examined. They included the effects of using different numbers of hidden units and the effects of using only position parameters in the cost function.

Marcelino Lazaro,et al[2] proposed a new cost function for supervised training of neural networks in binary classification applications. This cost function aimed at reducing the probability of classification error by reducing the overlap between distributions of the soft output for each class. The non-parametric Parzen window method, with Gaussian kernels, was used to estimate the distributions from the training data set. The cost function had been implemented in a GRBF neural network and had been tested in a motion detection application from low resolution infrared images, showing some advantages with respect to the conventional mean squared error cost function and also with respect to the support vector machine, a reference binary classifier.

Sven F. Crone[3] studied time series prediction in artificial neural networks which generally minimise a symmetric statistical error, such as the sum of squared errors, to learn relationships from the presented data. However, applications in business elucidate that real forecasting problems contain non-symmetric errors. The costs arising from suboptimal business decisions based on over-versus underprediction were dissimilar for errors of identical magnitude. To reflect this, a set of asymmetric cost functions was used as objective functions for neural network training, deriving superior forecasts even for white noise time series. Some experimental results were computed using a multilayer perceptron trained with various asymmetric cost functions, evaluating the performance in competition to conventional forecasting methods on a white noise time series extracted from the popular airline passenger data.

Sven F. Crone[4] studied time series prediction in artificial neural networks which generally minimizes a standard statistical error, such as the sum of squared errors, to learn relationships from the presented data. However, applications in business had shown that real forecasting problems require alternative error measures. Errors, identical in magnitude, caused different costs. To reflect this, a set of asymmetric cost functions was proposed as novel error functions for neural network training. Consequently, a neural network minimized an asymmetric cost function to derive forecasts considered preeminent regarding the original problem. Some experimental resulted in forecasting a stationary time series using a multilayer perceptron trained with a linear asymmetric cost function are computed and evaluated the performance in competition to basic forecast methods using various error measures.

S. Weaver, L. Baird, and M. Polycarpou[5] demonstrated that the slow learning of neural-network function approximators could be attributed to interference, which occured when learning in one area of the input space causes unlearning in another area. To mitigate the effect of unlearning, this paper developed an algorithm that adjusts the weights of an arbitrary, nonlinearly parameterized network such that the potential for future interference during learning was reduced. This was accomplished by the reduction of a objective cost function that combined the approximation error and a term that measures interference. Analysis of the algorithms convergence properties shown that learning reduced future unlearning. The algorithm could be used either during on-line learning or used to condition a network to have immunity from interference during a future learning stage. Simulations demonstrated how this new learning algorithm speeded training in various situations due to the extra cost function term.

B.Gunsel and C. Guzelig[6] proposed that the estimation of smoothing parameters is one of the difficult problems in using regularization techniques for image restoration. The objective of this paper was shown that Cellular Neural Networks (CNNs) incorporated with a learning algorithm can be useful in adaptive learning of smoothing parameters of regularization. Therefore, first a CNN model was designed to minimize a regularization cost function which was in quadratic form. The connection weights of this CNN were obtained by comparing the cost function with a Lyapunov function of the CNN. Unlike the common approaches in the literature, instead of learning connection weights of neural networks, they proposed supervised learning of the regularization smoothing parameters by a modified version of the Recurrent Perceptron Learning Algorithm (RPLA) which was recently developed for completely stable CNNs operating in a bipolar binary output mode. It was concluded that CNNs with the RPLA provides us to determine a set of suitable smoothing parameters resulting in a robust restoration of noisy images. For comparison purposes, experimental results obtained by median filter are also reported.

D.P. Mandic, and J. A. Chambers[7] addressed the choice of the coefficients in the cost function of a modular nested recurrent neural-network (RNN) architecture, known as the pipelined recurrent neural network (PRNN). Such a network could cope with the problem of vanishing gradient, experienced in prediction with RNNs. Constraints on the coefficients of the cost function, in the form of a vector norm, were considered. Unlike the previous cost function for the PRNN, which included a forgetting factor motivated by the recursive least squares (RLS) strategy, the proposed forms of cost function provide forgetting of the outputs of adjacent modules based upon the network architecture. Such an approach takes into account the number of modules in the PRNN, through the unit norm constraint on the coefficients of the cost function of the PRNN. This shown to be particularly suitable, since due to inherent nesting in the PRNN, every module gives its full contribution to the learning process, whereas the unit norm constrained cost function introduced a sense of forgetting in the memory management of the PRNN. The PRNN based upon a modified cost function outperforms existing PRNN schemes in the time series prediction simulations presented.