NEW ALGORITHM FOR IMPROVING PREDICTION PERFORMANCE IN MODIFIED RADIAL BASIS FUNCTION NETWORK

by

Lim Eng Aik
(1643912301)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

INSTITUT MATEMATIK KEJURUTERAAN
UNIVERSITI MALAYSIA PERLIS

2020

ACKNOWLEDGMENT

After a long way of writing, I completed this thesis. This thesis has received many helpful comments and suggestions from many people during the time of writing. I am indebted to my supervisor, Ts. Dr. Tan Wei Hong, and my co-supervisor, Assoc. Prof. Ts. Dr. Ahmad Kadri Junoh, for their useful suggestions and guidance during the research and the reports. Their influence through useful discussions is immeasurable. Sincere thanks to my parents and my wife for their love and encouragement that motivates me for writing this thesis. At last, thanks to my three energetic children, Lim Han Yang, Lim Han Feng, and Lim Jia Xuan, for their never-ending joys of activities that motivate me for keep-on going throughout the time I were writing this thesis, even though the tiredness felt like can sleep for days.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADALINE	Adaptive Linear Element
ANN	Artificial Neural Networks
AIC	Akaike Information Criterion
AUC	Area Under Curve
BOD	Biochemical Oxygen Demand
DWKM	Distance-Weighted K-Means
EM	Expectation Maximization
GD	Gradient Descent
IRBFN	Improved Radial Basis Function Network
KM	K-Means
LMS	Least Mean Squares
MATLAB	Matrix Laboratory
NN	Neural Networks
OLS	Orthogonal Least Squares
PSO	Particle Swarm Optimization
QEA	Quantum Evolutionary Algorithm
QRBFN	Quantum Evolutionary Radial Basis Function Network
RBFN	Radial Basis Function Networks
RMSE	Root Mean Squared Error

Algoritma Baru untuk Peningkatkan Prestasi Ramalan dalam Rangkaian Fungsi Asas Radial Terubahsuai

ABSTRAK

Dalam rangkaian saraf, ketepatan rangkaiannya bergantung kepada dua faktor kritikal, iaitu pusat dan nilai-nilai pemberat rangkaian. Rangkaian fungsi asas radial (RFAR) merupakan sejenis rangkaian ke-depan yang mampu melaksanakan taksiran tak linear pada set data yang tidak diketahui, klasifikasi, pengecaman corak, sistem kawalan, dan pemprosesan imej. Bagaimanapun, terdapat beberapa kelemahan rangkaian bagi RFAR seperti masa pengiraan yang lama untuk set data yang besar, algoritma kemaskini pemberat dan pemilihan pusat yang kurang tepat menyebabkan kejituan rendah. Titik data terhad atau titik data berlebihan boleh menjejaskan latihan RFAR. Maka, saiz yang bersesuaian diperlukan demi memastikan Rangkaian fungsi asas radial dilatih dengan saiz set data yang sesuai untuk mengurangkan masa pengiraan tanpa mempengaruhi ketepatannya secara mendadak. Untuk kemaskini pemberat RFAR, algoritma penurunan kecerunan mudah terperangkap dalam minima tempatan oleh pemberat rawak yang dijana pada peringkat awal latihan. Sementara itu, pemilihan pusat dengan algoritma K-min dikenali atas sensitiviti dan pergantungan yang tinggi kepada pemilihan pusat awal dari input set data. Maka, kerja ini mencadangkan penyelesaian bagi menangani kelemahan tersebut melalui pengubahsuaian pada beberapa bahagian algoritma RFAR untuk peningkatkan prestasi. Pertama, kerja ini mencadangkan formula pengurangan set data baru bagi mendapatkan bilangan set data yang sesuai untuk latihan rangkaian. Seterusnya, pengubahsuaian algoritma penurunan kecerunan dicadangkan untuk pengemaskinian berat RFAR semasa latihan. Kemudian, algoritma K-min pemberat berjarak dicadangkan bagi mendapatkan pusat permulaan yang lebih tepat untuk RFAR. Akhirnya, kerja ini mencadangkan model baru melalui gabungan algoritma evolusi kuantum (QEA) dan RFAR. RFAR yang dicadangkan ini menunjukkan kebolehannya dalam pencarian global dan pengoptimuman tempatan bagi memberikan ketepatan yang lebih baik dalam ramalan keputusan. Semua RFAR diubahsuai yang dicadangkan telah diuji terhadap RFAR standard dalam ketepatan ramalan ke atas empat model bukan linear dari kesusasteraan, dan empat dataset dunia nyata yang terdiri daripada dua dataset siri masa (dataset pencemar udara dan forex pasangan mata wang EURUSD), dan dua lagi data set ialah Oksigen Biokimia Dataset Permintaan (BOD) dan dataset pertumbuhan Phytoplankton. Rumusan pengurangan dataset yang dicadangkan dijalankan melalui eksperimen yang mana data diuji dengan pengurangan saiz langkah 5 peratus. Keputusan RFAR yang dicadangkan ini dibandingkan dengan nilai ralat punca min kuasa dua (RPMKD) dan keluasan bawah lengkung (KBL) dengan RFAR standard. Kes pengurangan dataset yang dicadangkan menghasilkan keputusan purata lebih dari 50 peratus penurunan penggunaan masa dan pengurangan 20 peratus dalam RPMKD. Sementara itu, RFAR yang dicadangkan menghasilkan keputusan yang lebih baik dan teguh dengan peratusan peningkatan purata lebih daripada 40 peratus dalam keputusan RPMKD dan KBL.

New Algorithm for Improving Prediction Performance in Modified Radial Basis Function Network

ABSTRACT

In neural networks, the accuracies of the networks are primarily relying on two critical factors, which are the centers and networks weight values. The feed-forward network known as Radial basis function network (RBFN) capable of performing nonlinear approximation on an unknown dataset, classification, pattern recognition, control system, and image processing. However, there are some disadvantages of the RBFN network, such as longer computation time for large datasets, less efficient weight updating, and center selection algorithms that cause low accuracy are identified. Limited data points or overload data points can affect the training of RBFN. Hence, proper size for dataset is required to ensure RBFN is trained using suitable dataset size to lessen the computational time without a significant influence on the accuracy. For RBFN weight updating, the gradient descent (GD) algorithm easily trapped in local minima by random weight generated during the initial stage of training. Meanwhile, the center's selection using the K-means algorithm is known for its sensitivity and high dependency to initial center selection from the input dataset. Therefore, this work proposed solutions for these mentioned disadvantages through modification on a few parts of the RBFN algorithm to improve their performance. First, this work proposed a new dataset reduction formula to obtain a suitable number of a dataset for network training. Next, a modified steepest descent algorithm was proposed for RBFN weight updating during training. Then, a new distance-weighted K-means algorithm is proposed for obtaining more accurate initial centers for RBFN. Finally, this work proposed a new model through a combination of quantum evolutionary algorithm (QEA) and RBFN known as QRBFN. This proposed RBFN demonstrated its abilities in global search and local optimization to effectively provide better accuracy in prediction results. All proposed modified RBFN was tested against the standard RBFN in predictions accuracy on four nonlinear models from literature, and four real-world datasets that consist two time-series datasets (Air pollutant dataset and forex pair EURUSD dataset), and other two datasets are Biochemical Oxygen Demand (BOD) dataset, and Phytoplankton growth dataset. The proposed dataset reduction formula was conducted through experiments where data was tested by a 5 percent step size reduction. The results of this proposed RBFN are compared for root mean square error (RMSE) and area under curve (AUC) values with standard RBFN. The proposed dataset reduction case yielded average results over a 50 percent decrease in time usage and a 20 percent reduction in RMSE. Meanwhile, all proposed RBFN yielded better results and robustness with an average improvement percentage of more than 40 percent in RMSE and AUC results.

CHAPTER 1 : INTRODUCTION

1.1 Neural Network

Artificial Neural Network (ANN) is an artificial neuro-processing model with neuron units that mimic the human biological brain in processing information and data. The heart of this paradigm is the unusual structure of information processing systems emphasizing parallel computation. The ANN architecture is built from a huge number of connected processing units known as neurons that functioning simultaneously to solve a specific problem (Haykin, 1994). The ANN, like the human brain, learn from the pattern of data and adapt from examples. Furthermore, an ANN can be constructed for a particular application, for example, pattern recognition, process control, function approximation, and data classification through learning from training data points that feed into the network. The learning of the network comprises the synaptic weight updating of the network to attain the objective of the design, whereas learning in actual biological neuro-system comprises adjustments of the synaptic connections between neurons.

ANN can be described as an artificial brain model that is constructed to resemble the actual brain executes a distinctive task of interest. Furthermore, ANN is typically implemented via electronic hardware or simulated in software on a computer. To achieve good performance, ANN must use a massive connection of processing units referred to as neurons.

Besides, ANN can be viewed as an "adaptive machine" designed to perform tasks by modeling the way the biological neuron system processes information, especially the brain according to Haykin (1994). Furthermore, ANN is notable for its ability to extract the meaning from complex or incomplete data are applicable in pattern extraction, trend detection, and data classification that is too complicated to be detected by either humans or other computer algorithms (Kubat, 1999).

1.2 Human Neurons and Artificial Neurons

There are abundant thing remains unknown regarding how precisely the brain training itself and process information, hence theories abound. In any human brain, a common signal is collected via and from others through a web of sheer architecture recognize as *dendrites*. The electrical pulses are sent out through a long thin stand called *axon* from a neuron that divided into numerous branches. By the end of each branch, the signal from the axon is turned into an electrical pulse by *synapse* that triggers by the activity, to all connected neurons. Hence, when the neurons acquire that excitatory input that is adequately large to distinguish from that restricts input, it then transmits an electrical pulse down to the axon. Learning takes place via altering the pulse strength in synapses, which impact one neuron to others from those changes (refer Figure 1.1). In 1943, McCulloch and Pitts (1943) took the first step towards the perceptron we use today by introducing a perceptron that mimicking the functionality of a biological neuron (McCulloch & Pitts, 1943). In 1949, Donald Hebb (1949) introduced Hebbian learning, which explains the simultaneous activation of cells that leads to pronounced increases in synaptic strength between those cells. Hebbian learning also provides a biological basis for errorless learning methods for education

and memory rehabilitation, which in the study of neural networks in cognitive function, it is often regarded as the neuronal basis of unsupervised learning (Hebb, 1949). Similar to biological axon function that transfer information, an artificial neuron is a mathematical function conceived as a model of biological neurons, which is an elementary unit in an artificial neural network. The artificial neuron receives one or more inputs that representing excitatory postsynaptic potentials at neural dendrites and sums them to produce an output that representing a neuron action potential that transmitted along its axon. The sum is passed through a nonlinear function known as an activation function or transfer function (Haykin, 1994).

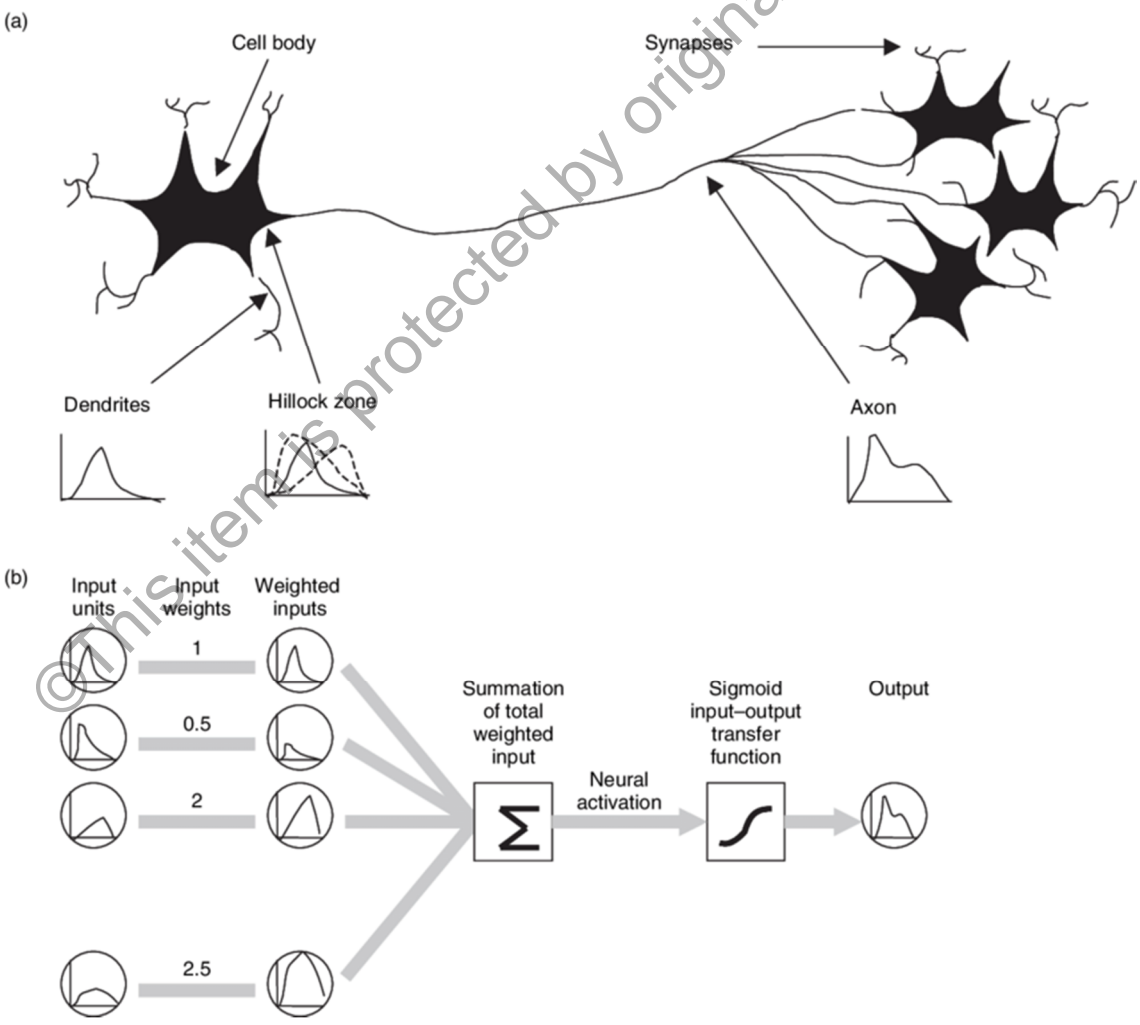

Figure 1.1. The Synapse. (a) The human synapse, and (b) artificial neural synapse (Recknagel, 2008)

1.3 Research Problems

The NN model has attracted significant criticisms from other researchers (Adam, Karras, Magoulas, & Vrahatis, 2014; Cao, Wang, Ming, & Gao, 2017; Kolasa, Długosz, Talaśka, & Pedrycz, 2018; Kusy & Kowalski, 2018; Lin, Xiang, Wang, & Yang, 2017; Lolli, 2017; Qiao, Li, & Li, 2016; Rao, Lu, Liu, & Xu, 2017; Samiee, Iosifidis, & Gabbouj, 2017; Seifollahi, Yearwood, & Ofoghi, 2012; Sodhi, Chandra, & Tanwar, 2014; Sun, Su, & Wang, 2018). Most criticisms involve the optimal number of training dataset, center selection, and networks weight, such as the following:

i) NN is strongly affected by the networks weight value, and most weight factor value problem often converges to local minima that lead to low accuracy results (Cao et al., 2017; Sodhi et al., 2014). Limited studies are found on improving the NN initial value problem in the predictive control field (Cao et al., 2017). Most of the recent solution reported in works of literature focuses on evolutionary algorithms in selecting optimal networks weight value (Cao et al., 2017; Kolasa et al., 2018). In addition, some of the works that use conjugate gradient in weight updating of the neural network, gain a significant improvement in results as reported in Zhang et al. (2019) that applied conjugate gradient in Wirtinger differential operator for faster convergence (Zhang, Liu, Cao, Wu, & Wang, 2019). Gurevich and Stuke (2020) use conjugate gradient in a multilayer neural network for prediction and curve-fitting purposes and gain good results in prediction (Gurevich & Stuke, 2020). However, limited work in solving network weight values problem using numerical approach (e.g., gradient

descent method or Steepest descent method) are reported (Cao et al., 2017). Some of the recent works using Steepest descent are found in Misev and Hills (2018) article that applied the steepest descent for optimizing Runge-Kutta coefficient for data prediction (Misev & Hills, 2018), Kumar and Rajasekhar (2017) performing a comparison of Levenberg-Marquardt and steepest descent algorithms, and the results are less significant improvement for data prediction results (Kumar & Rajasekhar, 2017). Additionally, the steepest descent method is considered as one of the simplest minimization methods for unconstrained optimization (Djordjevic, 2019; Muren, Wu, Zhou, Du, & Lv, 2019), and it also has a low computational cost with low matrix storage requirement for the computations of derivatives to be solved for the search direction compared to other search method such as evolutionary algorithms (Johari, Rivaie, & Mamat, 2018; Napitupulu, Sukono, Mohd, Hidayat, & Supian, 2018). The steepest descent method demonstrated its reliability and robustness for solving extremely ill-conditioned problems and yield significant improvement on results (Asl & Overton, 2020; Trang, Thoi, Khac, Chau, & Ao, 2019; Snyman, 2005; Zou & Magoules, 2019). Therefore, this thesis focuses on formulating a new weight updating algorithm by improving the existing steepest descent method for the input to the hidden layer of NN. Meanwhile, for hidden to the output layer of NN, this thesis also works on constructing a new quantum evolutionary algorithm integrates with a radial basis function network for optimizing the layer weights.

ii) Obtaining optimal center value for the hidden layer in NN during the training stage is vital for ensuring the high accuracy of the network. There is numerous approach to obtaining more accuracy center values reported in the literature in the past few years such as eigenvalues based center selection algorithm (Hu,

You, Liu, & He, 2018), replacing Gaussian function with categorial tuple function (Alex Alexandridis, Chondrodima, Giannopoulos, & Sarimveis, 2017), and modified K-means algorithm (Hanmin, Hao, & Qianting, 2016; Liu, Yin, & Chai, 2016; Xiong, Hua, Lv, & Li, 2017) on the nonlinear dataset. Clearly, from the works of literature mentioned, improvement has been made to change the Gaussian function, such as changing Gaussian function to kernel Gaussian function (Liu et al., 2016), and changing the distance calculation in Gaussian function with the average distance between data points (Xiong et al., 2017), where the enhancement in distance calculation is performed. However, most of the approaches mainly focusing on replacing the Gaussian function or introducing a supplementary method to support the center calculation process instead of modifying the Gaussian internal function to avoid the K-means algorithm run into a local minimum. This is because each step K-means algorithm aims at making the objective function decreases, if the initial cluster center is near to a local minimum, it will cause the local minimum problems (Hanmin et al., 2016). From works of literature review found in Section 2.4, the uses of distance as weightage in determining the importance of nearby center points were not considered. This is because the classical K-means algorithm focus on fixing the number of centers equal as a training data point (Du, Li, & Fei, 2010), or the center is randomly selected from training data as the center (Castro & Zuben, 2001; Mashor, 2000). While the recent improvement for the K-means algorithm focuses on joining the evolutionary algorithm for aiding to find a better center point (Cao et al., 2017; Kolasa et al., 2018). Hence, this thesis focuses on using distance as a weight for getting the optimal initial center for the networks and obtains better performance in prediction results. The reason

is that the proposed approach using distance as weight as part of the criteria for center selection requires less complicated algorithms compared to joining the evolutionary algorithm into the K-means algorithm for computation.

iii) The number of training dataset plays an important role in determining the prediction accuracy. According to the Ougiaroglou team (Ougiaroglou, Diamantaras, & Evangelidis, 2018) and, Bataineh and Marler (Bataineh & Marler, 2017) that conducting large data study, proof that dataset reduction was able to reduce training time and computation cost, especially in large dataset. Furthermore, Yousef and Kundu (Yousef & Kundu, 2014) demonstrated that the NN learning algorithm turns worse with an increasing dataset, as it may contain outlier data points within it. Yousef and Kundu (Yousef & Kundu, 2014) also stated that too little training dataset is also unsuitable for approximation use. Moreover, numerous research on data reduction uses optimization algorithm or other computational algorithms, for instances, support vector machine, random forest, optimum path forest, and nearest neighbour in determining the dataset that is important and separating the dataset that is less important (Chouvatut, Jindaluang, & Boonchieng, 2015; Kasemtaweechok, 2015; Mohsen, Kurban, Jenne, & Dalkilic, 2014; Ougiaroglou & Evangelidis, 2012; Juan & Iñesta, 2012; Shayegan & Aghabozorgi, 2014; Wang, Li, Liu, Zhang, & Zhang, 2014). In separated research conducted by Zhou et al. (2004) and Roy et al. (2008) showed that higher classification accuracy does not require large dataset (Roy, Leonard, & Roy, 2008); instead, NN only needs important dataset that can represent the overall shape of the case. Based on the mentioned works of literature work, all research mainly focus on an algorithm for reducing dataset

and selecting the suitable point from the dataset, none of these works of literature proposed any mathematical formula that can directly determine the suitable number of the dataset for training NN without a complicated algorithm. Hence, this thesis works on developing a new data reduction formula for determining a suitable number of dataset for NN training.

1.4 Research Objectives

i) To formulate a new data reduction formula, weight updating and K-means algorithms to improve the RBFN RMSE and AUC results.

ii) To construct a new quantum evolutionary algorithm that integrates into RBFN for optimizing input-to-hidden layer weights in RBFN networks weight.

iii) To compare the proposed RBFN with standard RBFN on its RMSE and AUC results for the nonlinear dataset simulation.

1.5 Important and Significant of the Research

Process control and prediction are vital to the automated operation and monitoring of complex technical processes that mostly behave in a nonlinear system. These processes represent the unit operations of automatic adjustment energy consumption, material transformation, and adjustment of the production line. Energy is employed to transform raw materials and create a useful product. The processes may, for example, include the production of electricity from the solar panel, the production,

and generation of ethanol through biomass; the parting of petrol from crude oil; or the waste-water treatment.

With intelligent control and prediction, processes are monitored, influenced, and adjusts as they happen. Without automation control and prediction on the system, disaster happens, for example, the Chernobyl incident (Labib, 2014) and Flixborough incident at Bhopal (Gehlawat, 2005; Macleod, 2014; Tinham, 2007). Furthermore, recent global industrialization, the Industry 4.0 (Bahrin, Othman, Azli, & Talib, 2016; Gilchrist, 2016; Lasi, Fettke, Kemper, Feld, & Hoffmann, 2014) has urge industries to set on the direction for increases in automation that will take place in all work sectors.

Hence, the need for a robust and accurate prediction simulator to perform control of the nonlinear system is extremely important to deal with the global trend in Industry 4.0. Basically, simulators are built on mathematical models of the NN, such as RBFN. Therefore, the progress of establishing a good prediction model plays a major role in obtaining good performance via the simulator. To be able to solve center selection, network weight, and obtaining optimal dataset problems that occur in the input and hidden layer will help in yielding more accurate and simulator.

The major beneficiaries of this research are manufacturing industries, telecommunication companies, energy corporations, financial institutions, oil and chemical industries. The beneficial aspects treated in this research focus on prediction accuracy, which can, in turn, provide accurate prediction and control over a nonlinear system and prevent the occurrence of disasters due to inaccuracy in prediction value. The output expected from the research is the modified RBFN model, for the purpose of

reducing process control failure, increase prediction accuracy, and prevent disasters from happening. It can be valuable for serving the industry and help open a new area of research.

1.6 Research Scope

The primary focus of this work is on the accuracy of the prediction results for NN. The focuses are restricted to study in training dataset size, hidden-layer networks weight, and input-layer centers selection of NN. Simulation of NN is performed using a different function from literature and real-world datasets that are highly nonlinear cases for comparison between standard RBFN and proposed modified RBFN for validation.

1.7 Organization of Thesis

The contributions made in this thesis are divided into parts, as shown in Figure 1.2:

i) The first part consists of the description of the NN, which is presented in this chapter, while the use of the NN model, along with its works of literature reviews, is presented in Chapter 2.

ii) In Chapter 3, the flow of obtaining the dataset reduction formula is explained. This chapter focused on reducing the training set size without inflicting a significant loss in approximation accuracy. The formula proposed in this chapter aims to reduce training dataset size and to maintain sufficient datasets which significantly interpret the form or dispersion of a model of the problem. This new formula able to calculate the suitable size of a training dataset for NN while

it does not cause a significant change in accuracy as the entire training set is applied. The proposed formula reduces the number of training dataset size by calculating the total training dataset size with Fibonacci retracement ratio and sums it with a bias.

Next, the steps involve in modifying the network's weight updating algorithm is explained. This modified RBFN that outperforms the standard RBFN in terms of accuracies using modified updating steps from gradient descent algorithm is presented.

To further improve the prediction accuracy, obtaining optimal center values are vital. Here, a modified version of the K-means algorithm using a new distance weightage function is introduced, formulated, and explained. This fast algorithm for training RBF networks that yield high accuracies is presented, where the initial input centers are selected through the Distance-Weighted K-Means algorithm.

Finally, the steps of constructing the quantum evolutionary algorithm are provided and explained at the end of Chapter 3. The proposed modified RBFN used quantum computing combined with an evolutionary algorithm to optimize the algorithm of RBFN.

iii) The simulation application and performance of results for all nonlinear function and the real-world dataset are presented and discussed in Chapter 4. The proposed networks efficiency is demonstrated through the application of eight experimental models, with four nonlinear models from literature, two real-world problems data obtained from Aik (2006) and two real-world time-series data from XM MetaTrader 4 trading platform (XM, 2018) and air pollutant dataset