UNIVERSITI PUTRA MALAYSIA

PREPROCESSING AND PRETRAINING OF MULTILAYER FEED FORWARD NEURAL NETWORK

ROYA ASADI

FSKTM 2009 10

PREPROCESSING AND PRETRAINING OF MULTILAYER FEED FORWARD NEURAL NETWORK

By

ROYA ASADI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master

December 2009

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

PREPROCESSING AND PRETRAINING OF MULTILAYER FEED FORWARD NEURAL NETWORK

By

ROYA ASADI

December 2009

Chairman: Norwati Mustapha, PhD

Faculty: Computer Science and Information Technology

The main problem for Supervised Multi-layer Neural Network (SMNN) model lies in finding the suitable weights during training in order to improve training time as well as achieve higher accuracy. The important issue in the training process of the existing SMNN model is initialization of the weights. However, this process is random and creates the paradox of low accuracy and high training time.

In this study, a Multi-layer Feed Forward Neural Network (MFFNN) model for classification problem is proposed. It consists of a new preprocessing technique which combines data preprocessing and pre-training that offer a number of advantages; training cycle, gradient of mean square error function, and updating weights are not needed in this model. The proposed technique is Weight Linear Analysis (WLA) based on mathematical,

statistical and physical principles for generating real weights by using input values. WLA applies global mean and vectors torque formula to solve the problem. We perform data preprocessing for generating normalized input values and then applying them by a pre-training technique in order to obtain the real weights. The normalized input values and real weights are applied to the MFFNN model in one epoch without training cycle. In MFFNN model, thresholds of training set and test set are computed by using input values and real weights. In training set each instance has one special threshold and class label. In test set the threshold of each instance will be compared with the range of thresholds of training set and the class label of each instance will be predicted.

To evaluate the performance of the proposed MFFNN model, a series of experiment on XOR problem and two datasets, which are SPECT Heart and SPECTF Heart was implemented. As quoted by literature, these two datasets are difficult for classification and most of the conventional methods do not process well on these datasets. For experiment result, Standard Back Propagation Network (BPN) as SMNN model is considered. SBPN is changed to MFFNN model by using WLA technique. Accuracy of MFFNN model using WLA is compared with several strong classification models and SBPN using best and latest pre-training techniques. Our results, however, show that the proposed model has been given high accuracy in one epoch without training cycle. The accuracies of 94% for SPECTF Heart and 92% for SPECT Heart which are the best results.

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

PRA-PEMPROSESAN DAN PRA-LATIHAN BAGI RANGKAIAN NEURAL MULTI-ARAS SUAP HADAPAN

Oleh

ROYA ASADI

Disember 2009

Pengerusi: **Norwati Mustapha, PhD**

Fakulti: **Sains Komputer dan Teknologi Maklumat**

Masalah utama bagi Rangkaian Neural Multi-aras Terselia (SMNN) adalah dalam mencari pemberat yang bersesuaian semasa latihan, bagi meningkatkan masa latihan dan mencapai ketepatan yang tinggi. Isu terpenting dalam latihan SMNN adalah mencari nilai awal pemberat. Walau bagaimanapun, proses ini adalah rawak dan mencetus paradoks iaitu ketepatan yang rendah dan masa latihan yang tinggi.

Dalam kajian ini, sebuah model Rangkaian Neural Multi-aras Suap Hadapan untuk masalah pengkelasan adalah dicadangkan. Ianya mengandungi teknik baru pra-pemprosesan yang menggabungkan pra-pemprosesan data dan pra-latihan yang menawarkan beberapa kebaikan; kitaran latihan, fungsi kecerunan ralat min kuasa dua, dan pengemaskinian pemberat adalah tidak lagi diperlukan. Teknik yang telah dicadangkan ialah *Weight Linear Analysis* (WLA) yang berdasarkan prinsip-prinsip matematik, statistik

dan fizik bagi menjana pemberat sebenar dengan menggunakan nilai-nilai input. WLA menggunakan purata global dan formula *vectors torque* untuk menyelesaikan masalah tersebut. Kami melaksanakan pra-pemprosesan untuk menjana nilai input ternormal yang kemudiannya digunakan oleh teknik pra-latihan dalam tujuan untuk mendapatkan pemberat sebenar. Nilai-nilai input ternormal dan pemberat sebenar digunakan pada model MFFNN dalam satu epoksi tanpa kitaran latihan. Dalam model MFFNN, nilai ambang bagi set latihan dan set ujian dikira dengan menggunakan nilai input dan pemberat sebenar. Dalam set latihan, setiap instan mempunyai satu nilai ambang dan label kelas tersendiri. Dalam set ujian, nilai ambang bagi setiap instan akan dibandingkan dengan julat nilai ambang bagi set latihan dan label kelas bagi setiap instan akan diramalkan.

Bagi menilai prestasi model MFFNN yang telah dicadangkan, beberapa siri eksperimen ke atas masalah XOR dan dua set data, iaitu SPECT Heart dan SPECTF Heart telah dilaksanakan. Sebagaimana dinyatakan dalam kajian literatur, kedua-dua set data ini adalah sukar bagi pengkelasan dan kaedah konvensional tidak mendapat hasil pemprosesan yang baik daripada kedua-dua set data ini. Bagi hasil eksperimen, Standard Back Propagation Network (BPN) sebagai model SMNN adalah di ambil kira. SBPN ditukarkan kepada model MFFNN dengan menggunakan teknik WLA. Ketepatan model MFFNN menggunakan WLA dibandingkan dengan beberapa model pengkelasan yang baik dan model SBPN menggunakan teknik-teknik pra-latihan yang terbaik dan terkini. Hasil kami, walau bagaimanapun, menunjukkan model yang dicadangkan telah memberi ketepatan yang tinggi dalam satu epoksi tanpa kitaran latihan. Ketepatannya adalah 94% untuk SPECTF Heart dan 92% untuk SPECT Heart iaitu hasil yang terbaik.

ACKNOWLEDGEMENTS

First of all, thank to God, the most Gracious and most Merciful.

Many people have made contributions in completing this work. My deepest appreciation and gratitude to the supervisory committee led by Dr. Norwati Mustapha, as well as the committee members; Assoc. Prof. Dr. Md. Nasir Sulaiman and Assoc. Prof. Dr. Nematollah Shiri Selamat for their virtuosity, ideas, intellectual experiences, and support that paved the way to this research work. Also, my most affectionate gratitude to my beloved parents, sisters, and brother.

Finally, sincere thanks towards everyone in the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia for their intellectual and social contributions, especially for making my research years to be most memorable and enjoyable times.

APPROVAL

I certify that a Thesis Examination Committee has met on 2th December 2009 to conduct the final examination of Roya Asadi on her Master thesis entitled "PREPROCESSING AND PRETRAINING OF MULTILAYER FEED FORWARD NEURAL NETWORK" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The committee recommends that the student be awarded the Master degree.

Members of the Thesis Examination Committee were as follows:

Azmi bin Jaafar, PhD.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Ramlan Mahmod, PhD.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Ali Mamat, PhD.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Naomie Salim, PhD.
Associate Professor
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
Malaysia
(External Examiner)

BUJANG BIN KIM HUAT, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 15 January 2010

Saya mengesahkan bahawa satu Jawatankuasa Peperiksaan Tesis telah berjumpa pada 2 Disember 2009 untuk menjalankan peperiksaan akhir bagi Roya Asadi bagi menilai tesis beliau yang bertajuk PRA-PEMPROSESAN DAN PRA-LATIHAN BAGI RANGKAIAN NEURAL MULTI-ARAS SUAP HADAPAN mengikut Akta Universiti dan Kolej Universiti 1971 dan Perlembagaan Universiti Putra Malaysia [P.U.(A) 106] 15 Mac 1998. Jawatankuasa tersebut telah memperakukan bahawa calon ini layak dianugerahi ijazah Master.

Ahli Jawatankuasa Peperiksaan Tesis adalah seperti berikut:

Azmi bin Jaafar, PhD.
Profesor Madya
Fakulti Sains Komputer dan Teknologi Maklumat
Universiti Putra Malaysia
(Pengerusi)

Ramlan Mahmod, PhD.
Profesor Madya
Fakulti Sains Komputer dan Teknologi Maklumat
Universiti Putra Malaysia
(Pemeriksa Dalam)

Ali Mamat, PhD.
Profesor Madya
Fakulti Sains Komputer dan Teknologi Maklumat
Universiti Putra Malaysia
(Pemeriksa Dalam)

Naomie Salim, PhD.
Profesor Madya
Fakulti Sains Komputer dan Sistem Maklumat
Universiti Teknologi Malaysia
Malaysia
(Pemeriksa Luar)

BUJANG BIN KIM HUAT, PhD
Profesor dan Timbalan Dekan
Sekolah Pengajian Siswazah
Universiti Putra Malaysia

Tarikh: 15 Januari 2010

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master. The members of the Supervisory Committee were as follows:

Norwati Mustapha, PhD.
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairperson)

Md. Nasir Sulaiman, PhD.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Nematollah Shiri, PhD.
Associate Professor
Department of Computer Science
Concordia University in Canada, Quebec
(Member)

HASANAH MOHD GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 11 February 2010

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institutions.

ROYA ASADI

Date: 22 Jun 2009

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADALINE	ADAptive Linear Neuron
ART	Adaptive Resonance Theory
BPN	Back-Propagation Network
CLIP	Cover Learning using Integer Programming
DPT	Delta Pre-Training
HDs	Hamming Distances
K-NN	K-Nearest Neighbor
LMS	Least Mean Square algorithm
LVQ	Learning Vector Quantization
MLP	Multi-layer Perceptron
MSE	Mean Square Error
PCA	Principal Component Analysis
PNP	Paralyzed Neuron Percentage
RBM	Restricted Boltzmann Machines
RN	Recurrent Networks
SBPN	Standard Back Propagation Network
SCAWI	Statistically Controlled Activation Weight Initialization
SLP	Single-Layer Perceptron
SOM	Self-organizing Map
SPECT	Single Proton Emission Computed Tomography
SMNN	Supervised Multi-layer Neural Network
MFFNN	Supervised Multi-layer Feed Forward Neural Network

TLU	Threshold Logic Unit
VLSI	Very Large Scale Integrated
WLA	Weights Linear Analysis
XOR	Exclusive OR

CHAPTER 1

INTRODUCTION

1.1 Background

Neural networks are suitable for extracting rules, quantitative evaluation of these rules, clustering, self-organization, classification, regression feature evaluation, and dimensionality reduction (Han and Kamber, 2001; Mitra et al., 2002). Back-propagation network (BPN) is the best example of a parametric method for training supervised multi-layer perception neural network for classification. BPN has the ability to learn biases and weights (Han and Kamber, 2001). It is a powerful method to control or classify systems that use data to adjust the network weights and thresholds for minimizing the error in its predictions on the training set. Learning in BPN employs gradient-based optimization method in two basic steps: to calculate the gradient of error function and to compute output by the gradient. Also, BPN compares each output value with its sigmoid function in the input forward and computes its error in BPN backward. This is considerably slow because biases and weights have to be updated in each epoch of learning (Craven and Shavlik, 1997).

Learning is the important property of neural networks. Neural networks are able to dynamically learn types of input information based on their weights and properties. During learning, the weight of each value in hidden layers will be considered. As the domain becomes smaller and smaller, suitable weights will be obtained through this series of repeated trial and errors after several epochs. Suitable data pre-processing techniques are necessary to find input values while pre-training techniques to find desirable weights that in turn will reduce the training process. This is the essence of Supervised Multi-layer Neural Network (SMNN). Combination of data pre-processing and pre-training in SMNN results in worthy input values, desirable process, and higher performance in both speed and accuracy.

1.2 Problem Statement

In classification problem, Supervised Multi-layer Neural Network (SMNN) model such as Back-propagation Network (BPN) obtain its data by learning from the real-world environment. It also dynamically recognizes the type of input information through their weights and properties (Han and Kamber, 2001). Learning of SMNN models such as BPN is considerably slow because biases and weights have to be updated in each epoch during learning (Craven and Shavlik, 1997). Without pre-processing, the training for classification may be very slow and may not even complete.

Pre-processing consists of additional steps applied to help improve the accuracy, speed, and scalability during classification (Han and Kamber, 2001). Currently, data pre-processing and pre-training are the contributing factors in developing efficient techniques for fast SMNN processing at high accuracy and reduced training time (Van der Maaten et al., 2008; Hinton and Salakhutdinov, 2006). Nonetheless, the improvement in performance and results is not without cost. Finding the suitable input values and weights has become critical to at least maintain the same processing time and high accuracy results (Demuth et al., 2007; Andonie and Kovalerchuk, 2004; Jolliffe, 2002; Han and Kamber, 2001). Suitable input values and weights are extremely necessary because neural network lies on the foundation of “garbage-in, garbage-out”. Kim and Ra (1991) introduced a minimum bound for initialization of weight and Fernández-Redondo and Hernández-Espinosa (2001) proposed upper bound 0.1 plus lower bound. Using random number for weight initialization is disadvantage of this technique. Drago and Ridella (1992) introduced a method called Statistically Controlled Activation Weight Initialization (SCAWI). They used the meaning of paralyzed neuron percentage (PNP) and conceptualized on testing the number of times a neuron lies in a completed situation with acceptable error. Fernandez-Redondo and Hernandez-Espinosa (2000) and Funahash (1989) improved this method. However, SCAWI is also using random numbers; hence it is critical during training. Li et al., (1993) explained Delta Pre-Training (DPT). The core of DPT is using Delta rule instead of using random numbers, after this phase, SMNN model training process is carried out to complete network training. The weights are initialized with zero values by using Delta rule. Disadvantage of this technique is the initialization zero value that is not based on desired and real weights. Shimodaira (1994) introduced one pre-training technique based on geometrical considerations. Disadvantage of this technique the same DPT technique is the initialization zero value. Ho-Sub et al., (1995) classified the input values in three groups, whereby weights of the most important input are initialized with [0.5, 1], weights with the least important input are initialized with [0, 0.5], and the rest are initialized with [0, 1]. Weight initialization is at random yet. Keeni et al. (1999) introduced the idea for initializing weight range within the domain of [−0.77; 0.77]. The experiment achieved best mean performance for multi-layer perceptrons with only one hidden layer. Already, Weight initialization is at random. Zhang et al. (2004) and Fernández-Redondo and Hernández-Espinosa (2001) discussed several initial weighting methods in Min and Max, initial weights considered in domain of (-a, +a) are computed. Currently, Min and Max technique using standard BPN has an initial random weight in domain [-0.05, 0.05] (Fernández-Redondo and Hernández-Espinosa, 2001). Nonetheless, the disadvantage of Min and Max method is the initialization of random numbers that are critical during training. Latest and strong pre-training technique was introduced by Van der Maaten et al., (2008); Hinton and Salakhutdinov, (2006); DeMers and Cottrell, (1993) which is multi-layer auto-encoder networks. The feed forward neural network trains to minimize the mean squared error between the input and output by using sigmoid function. High-dimensional matrix may be reduced into low-dimensional matrix through extraction of node values in the middle hidden layer. In addition, auto-encoder/auto-Associative neural networks are neural networks that are trained to recall their inputs. When the neural network uses linear neuron and activation functions, auto-encoder processes are similar to PCA (Lanckriet et al., 2004). BPN advances global fine-tuning phase through auto-encoder to fine-tune the weights for optimization. The main disadvantages of this method are using random number for weight initialization and due to the high number of multi-layer auto-encoders connections in BPN training process, resulting in slow performance.

Therefore, current pre-training techniques apply random values for initial weights to reduce the training process (Van der Maaten et al., 2008; DeMers and Cottrell, 1993) but this approach resulted in a contradicting paradox between the accuracy result and training time (Zhang et al., 2004; Fernández-Redondo and Hernández-Espinosa, 2001).

In this study, a Multi-layer Feed Forward Neural Network (MFFNN) model for classification problem is proposed. It consists of a new preprocessing technique, called Weight Linear Analysis (WLA) which combines data preprocessing and pre-training that offer a number of advantages; training cycle, gradient of mean square error function, and updating weights are not needed in this model. WLA applies global mean and vectors torque formula to solve the problem. We perform data preprocessing for generating normalized input values and then applying them by new pre-training technique in order to obtain the real weights. The normalized input values and real weights are applied to the MFFNN model in one epoch without training cycle.

1.3 Objectives of research

The objectives of this research are as follows:

- To propose a Multi-layer Feed Forward Neural Network (MFFNN) model that increase classification accuracy and improve training time in a single epoch classification.

- To propose a new pre-training technique that combines with data pre-processing to generate real weights through the use of normalized input values.

1.4 Scope of research

This research focuses on combination of data pre-processing and pre-training techniques using weights linear analysis to increase classification accuracy and to reduce training time in the Multi-layer Feed Forward Neural Network (MFFNN) model. For experimental result, Standard Back Propagation Network is considered as Multi-layer feed forward models. All experiments are carried out under the domain of Exclusive-OR (XOR) problem using two datasets, which are SPECT Heart and SPECTF Heart. Training is performed through *F*-measure function with 10 folds of testset and is compared with results from several strong models. These datasets are suitable as benchmarking in multi-layer networks based on UCI Repository of Machine Learning and comparison (Kim and Zhang, 2007).

1.5 Research methodology

The research follows combination of data pre-processing and pre-training techniques to reduce training process in Multi-layer Feed Forward Neural Network (MFFNN) model at high speed and accuracy. The techniques apply mathematical, statistical, and physical principles for generating real weights by using normalized input values, which are essentially the output of data pre-processing phase.

The proposed technique, called Weights Linear Analysis (WLA), addresses the problem in two phases. During the first phase, WLA considers data pre-processing through vertical evaluation on input values matrix for generating normalized input values. The output of the first phase, which are normalized input values, are then used to compute weights of attributes. The proposed MFFNN, in turn, uses the output of WLA, which are normalized values and weights to classify the dataset. This approach resulted in a reduced training time because learning does not require computing mean square errors and updating weights in any training cycle.

Currently, pre-training is using multilayer encoders as feed-forward neural networks with odd number of hidden layers Van der Maaten et al. (2008) and DeMers and Cottrell (1993). This pre-training method uses initial random weights as opposed to the real weights used in the proposed method. As for the experimental setup, an Exclusive-OR (XOR) problem with two sets of dataset is chosen to illustrate the strength of the proposed technique. The first dataset, SPECTF Heart is a multivariate integer dataset while the second dataset, SPECT Heart is a multivariate categorical-binary dataset.

1.6 Contribution of research

Major contributions described in this thesis are listed below:

- Multi-layer Feed Forward Neural Network (MFFNN) model for single epoch classification, which does not require any training cycle, computation of gradient mean square error function, and updating weights. The approach uses Weights Linear Analysis (WLA) to combine data pre-processing and pre-training.