IOSUD - Universitatea Politehnica Timişoara
Şcoala Doctorală de Studii Inginereşti

Contributions to the Determination of Neural Network Architectures

PhD thesis - Summary

To obtain the scientific title of doctor at
Polytechnic University of Timisoara
in the field of Artificial Intelligence and Machine Learning

author Mohamed Lafif Tej
scientific leader Prof.univ.dr.ing. Stefan Holban

Abstract

This thesis aims to create a general method to determine the optimal neural network architecture using data mining techniques. Data mining are used to explore training data of a neural network to identify patterns and establish relationships that can lead to reasonable predictions of the optimal architecture of a multi-layer neural network.

There are no well-developed formal methods for a priori determination of the optimal architecture of neural networks. Most currently used methods are very limited and time-consuming. Successfully methods solve a few applications under specified conditions. Currently, we are far from getting a viable theory for determining the architecture of a neural network, depending on the complexity of the concerning problem to be solved.

The aim of this study intends to do is to determine a neural network architecture and the number of hidden layers and number of neurons in these layers. I intend in this respect to use Data Mining techniques to analyze the data to be processed by a neural network. The data to be processed by a multilayer neural network, I intend to analyze them through clustering techniques in the context of the way to work unsupervised and to correlate the number of groups obtained with the optimal number of hidden layers for a multilayer neural network. In determining the number of neurons on each hidden layer will use the results by clustering forms from the database analyzed on the basis of a reference range.

This thesis presents a new general method to determine the optimal neural network architecture using Linear Regression. Starting from clustering the dataset used to train a neural network it is possible to define Multiple Linear Regression models to determine the architecture of a neural network. This method is more flexible with different datasets types unlike other methods. The proposed method adapts to the complexity of training datasets to provide the best results regardless of the size and type of dataset.

The proposed method reduces the time allocated for network design. This method can make the design simple and easy and possible for a Non-specialist designer. The focus was to develop a fast solution (in terms of learning iterations) maintaining also an acceptable efficiency.

Keywords: Artificial intelligence, machine learning, neural networks architecture, data mining, clustering methods, multi-layer neural network, pattern recognition, regression analysis.

1. Introduction

The concerns of researchers were concentrated on the theory that computers can learn without being programmed to perform specific tasks [1], researchers interested in artificial intelligence [2] wanted to see if computers could learn from data. Machine learning gives computers the ability to learn without being explicitly programmed [3].

Artificial Neural Networks is Biologically Inspired [4]. A biologically inspired programming paradigm enables a computer to learn from observational data. More than two decades ago, neural networks were widely seen as the next generation of computing, one that would finally allow computers to think for themselves.

Achievement of optimum neural network architecture is necessary due to neural networks applications in medical imaging (localization of tumors and other pathologies, measuring tissue volume, Computer-assisted surgery, diagnosis, treatment planning), location of objects in satellite images, facial recognition, iris recognition, fingerprint recognition, traffic control systems and more [5]. These applications recommended as the current one in which have invested effort to satisfactory results as close to the ideal.

1.1.Problem Statement

In this thesis, we will present a scientific criticism of Neural Network architecture where there are no generally available analytical methods for determining the optimal number of hidden layers [6] and the number of neurons in these layers [7] to a multilayer neural network [8]. In the literature there is described a series of neural network architectures, but successfully solves a few applications under specified conditions. The design of the structure of the neural network is an extremely active area of research and does not yet have many definitive guiding theoretical principles.

In experiments made by various researchers specified for specific issues such as the optimal architecture, but they give results only under certain conditions. Choosing the number of hidden layers and the number of neurons is a complex issue and a critical step in the design of a neural network, for which there are no generally valid theories. Currently, we are far from having a viable theory for determining the size of a neural network, depending on the complexity of the problem to be solved [9].

1.2.Research Approach

I intend in this respect to use Data Mining techniques to analyze the data used to train a neural network. This method will work unsupervised to analyze the training data through clustering techniques and to correlate the number of groups obtained with the optimal number of hidden layers for a multilayer neural network. In determining the number of neurons on each hidden layer we will use the results by clustering training forms from the analyzed database on the basis of a reference distance.

In order to calculate the number of neurons in the hidden layers, a Multiple Linear Regression model using parameters obtained from clustering the training dataset will be developed. Use of this method allows the design of an optimal neural network to be unsupervised and will decrease its build time.

1.3.Outline of the thesis

The rest of this thesis is organized as follows:

- **Chapter 2** contains an analysis of the structures of Neural Networks and explanation of the effect of the architecture of neural network on the ability of learning of neural network. In addition, a presentation of the methods currently used to determine the architecture of the neural network.

- **Chapter 3** describes the method used to determine the optimal Neural Network Architecture using clustering techniques by presenting the stages, which we pass through to determine the architecture of the neural network. In addition, a discussion about clustering techniques used and a comparative study of clustering distance measures used to Determine Neural Network Architectures.
- **Chapter 4** describes another technique used to determine the optimal Neural Network Architecture by means of regression methods. The results obtained from clustering the training data are found to be useful to develop a Multiple Linear Regression model in accordance with the purpose of determining the number of hidden layers and the number of neurons on each layer for a multilayer neural network. A discussion on factors selected and the influence of each factor on the number of hidden layers and the number of the hidden neuron will be presented in this chapter.
- **Chapter 5** describes the importance of training data analysis to improve generalization capabilities of the Neural Network Architectures. It is shown that the generalization performance of neural networks is affected by the structure of the network.
- **Chapter 6** describes a comparison of the proposed method with the most currently used methods. This chapter shows how the proposed method performs well for the different type of datasets and how it’s adapted to the complexity of the training data to provide the best results regardless the size and type of dataset in contrast to other methods.
- **Chapter 7** presents a summary of the work presented in this thesis and concludes this work.

2. Structure of the Neural Network

The most important theoretical problem presented in this thesis is the correlation between the structure of the neural network and the learning ability of the network between the structure of the neural network and the learning ability of the network. To solve a complex problem using the neural network it needs a complex network structure represented by the number of hidden layers and the number of hidden units. The size of the network affects the learning capability of the neural network, which makes adding more hidden layers and neurons in these layers essential when the complexity of the problem to be solved increases.

Theoretically, it can be proved that a neural network with one advantage of using multiple hidden layers is improving the predictive ability of the network. The problem is how many hidden layers we need based on the level of complexity of the problem to be solved by the neural network without exceeding the required number of hidden layers to avoid the decrease of the accuracy in the test set. Increasing the number of hidden layers much more than required will cause the network to overfit to the training set that means it will learn the training data, but it won't be able to generalize to new unseen data.

The number of hidden neurons represent a critical issue in the design of a Neural Network, most currently used methods are Trial and Error, evolutionary algorithms, exhaustive search, and Growing and Pruning algorithms, but they are very limited and time-consuming.

In this chapter, the structures of Neural Networks will be analyzed and an explanation of the effect of the architecture of neural network on the ability of learning of neural network. In addition to the number of hidden layers and the number of hidden neurons, the structure of a neural network depends on many parameters represented by Cost function, Activation function, and Hyper-parameters [10].

Conclusions:

In this chapter, we represented the structure of the neural network and the effect of number of hidden layers and the number of hidden neurons on the learning ability of the network. In general, the network size affects the network complexity, but most importantly, it affects the generalization capabilities of the network.

Several widely used methods to determine the architecture of the neural network was discussed and the weaknesses of these methods are presented. It concluded that until now there is no general method to determine the structure of a neural network based on the complexity of the problem to be solved.

The training time and the generalization of the network is the main problem, which confronts us in the process of the design of the neural network structure. The number of hidden layers depends on the complexity of the problem to achieve good classification accuracy. A high number of the hidden layers may cause overfit to the training dataset.

3. Determination of optimal Neural Network Architecture using clustering techniques

This chapter presents a new method to determine the optimal Multi-layer Perceptron structure using pattern recognition and data mining. Using Clustering techniques on the dataset used to train the neural network and based on given criteria, we can define a number of clusters. The results obtained from clustering of training dataset can be used as an indicator to determine the level of complexity of the problem to be solved. Based on the number of clusters obtained we can determine the number of hidden layers for a multi-layer perceptron (MLP). This study gives more generalization capabilities to the proposed method.

Use of this method avoids making a complicated neural network structure by setting a number of hidden layers more than required. On the other hand, this method avoids using a few numbers of hidden layers which make the network unable to achieve satisfactory performance. Concerning the number of hidden neurons this method avoids using a large number of hidden neurons, because with a number of neurons more than required the network will memories the patterns instead of learning from the training set [11], which affects the generalization ability of the neural network to interpolate and extrapolate data that it has not seen before.

The idea behind this clustering method is to cluster the dataset used to train the neural network using methods of pattern recognition [3]. Thereby we obtain a set of homogeneous classes; in addition, we obtain a number of elements that do not belong to any cluster. Based on the information collected from clustering the training dataset we conclude that the number of clusters obtained in the case of at least 90% of input forms clustered is equal to the best possible number of hidden layers for a multi-layer neural network. The number of clusters must be taken as few as possible in order to obtain a network with the lowest number possible of hidden layers to decrease the complexity of the network.

Taking into consideration the stability of the clustering operation, including its meaning by increasing the value of Reference Distance the number of clusters obtained dose not vary provided that the extreme cases not considered. Such as a very short value of Reference Distance for which each element represents a cluster or relatively large value of Reference Distances for which all the elements placed in one cluster. The Reference Distance is where to cut the dendrogram obtained using hierarchical clustering to form clusters.

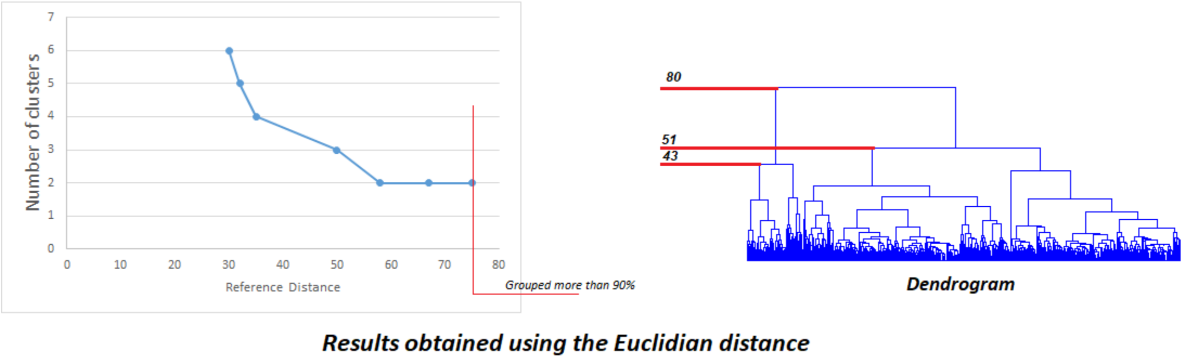

Figure 1: The relationship between the number of groups that can be obtained according to the reference distance (Euclidean or Manhattan) and corresponding Dendrogram obtained from the analysis of Landsat Satellite images dataset.

At this stage, we take into consideration linkage distances, such as complete-linkage clustering, single-linkage clustering, mean distance clustering, and the percentage of input forms that are clustered below and above the average distance. These parameters are considered in the step of designing the neural network. The number of clusters obtained using the proposed method is going to be considered the optimal number of hidden layers of the multi-layer neural network. Figure 2 explain the method of determining the optimal MLP structure.

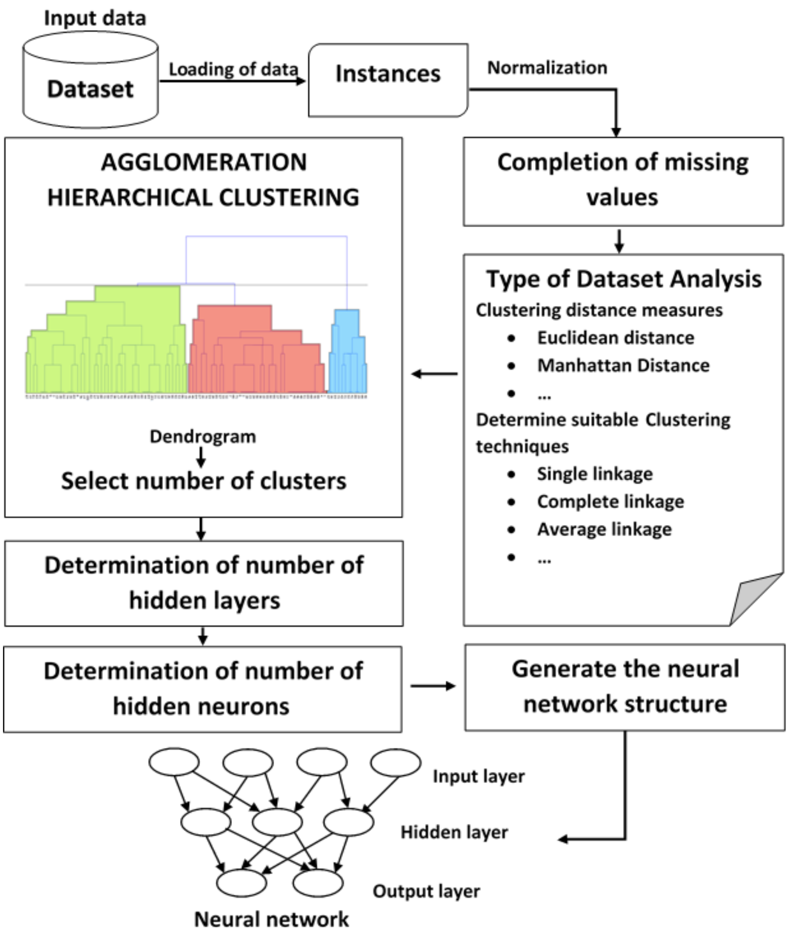

Figure 2. Proposed framework

After preparation of the training dataset by the elimination of noise, incomplete records and those records showing large dissimilarities to other data, a data analysis is necessary to determine the distance measure suitable with the type of data used. Implementation of agglomerative hierarchical clustering algorithm generates a Dendrogram, which will then help us to define the number of clusters, which will also help to determine the number of hidden layers.

3.1.Stages of the Method Used

The proposed method seeking to evaluate the level of complexity of the concerning problem by clustering the training data of a neural network and then interpret the obtained cluster to define the number of hidden layers. By clustering the dataset used to train the neural network based on conventional methods of pattern recognition [12] [13] following a set of criteria to generate a number of clusters. In this case, we can take the number of clusters obtained as the optimal number of hidden layers for the MLP structure [14]. This method will take into consideration several criteria which be discussed in this chapter. The algorithm used to design the neural network is developed through several stages.

The paragraph will be structured to highlight the stages of the proposed method. First stage is establishing the training dataset upon which the neural network learning will be done. Next stage is establishing the number of input neurons which will be equal to the number of features. Then determining the number of output neurons which will be equal to the number of classes if we have a classification problem or the number of output features vector in other cases. Then the stage of determining the number of hidden layers is based on clustering the training dataset according to a set of criteria in order to obtain a number of clusters equal to the optimal number of hidden layers to do this it is necessary to meet the following condition:

- The clustering of training dataset must be performed with a covering of at least 90% of the elements of the dataset. Because with at least 90% of training dataset elements we have adequate representation of the dataset forms and the result can be extrapolated to the entire dataset.
- The number of clusters obtained must be constant if we increase the value of reference distance, which indicates that the grouping is stable. The extreme cases are not taken in to account among them a very short value of reference distance in which each element of the training dataset is considered as a cluster, or a relatively large value of reference distance in which all elements of the training dataset are clustered into one single cluster.

By applying, the criteria described above the number of groups obtained will be considered as the optimal number of hidden layers for an MLP. The main aim of this study is to attain the best criteria, which lead to getting a number of clusters equal to the optimal number of hidden layers.

Based on experiment results it was observed that the number of clusters obtained by clustering the training dataset based on the criteria described above is approaching to be equal to the optimal number of hidden layers.

The point of this research is to determine the suitable criteria to correlate the number of the obtained cluster with the optimal number of hidden layers, which requires a comparative study of clustering distance measures [15]. At this point, a set of linkage distances are taken into consideration, such as complete-linkage clustering, single-linkage clustering, mean distance clustering, and the percentage of input forms that are clustered below and above the average distance.

3.2.Data mining techniques used

Data mining techniques are used to analyses datasets, extract useful information, and establish relationships based on pattern discovered from data to find solutions for some problems. Based on information extracted from data using data mining we can make reasonable predictions concerning relationships between reviled patterns in data.
In this chapter, several data mining techniques are used to analyze the training dataset to define the optimal MLP architecture, among them clustering, Regression method, classification algorithms, and prediction.
The clustering method helps to determine the optimal structure of a neural network depending on the results obtained through clustering the training dataset of a neural network using Agglomerative Hierarchical Clustering algorithm.
The Agglomerating Hierarchical Clustering algorithm helps to analyze data to create clusters based on the information obtained from data. Based on the relationship between elements of a dataset and the information extracted from each element the data are grouped into clusters.

Conclusions:

Using the proposed method, it is possible to determine the architecture of the neural network based on the complexity level of the problem to be solved.
Using clustering techniques, it is possible to highlight several common characteristics of input forms, on which they (input forms) can be classified into groups (clusters). The clustering of input forms generates a number of useful factors, which help to determine the optimal neural network architecture.
Based on the results obtained in this chapter it is observed that clustering algorithms used are affected by several factors which influence on the obtained number of hidden layers using the proposed method. The most important factor is the Reference Distance, the accuracy of the value of this factor is necessary to achieve the optimum number of hidden layers.
The comparative study of clustering distance measures lends on the proposed method more effectiveness and accuracy through the perfect selection of the appropriate distance measures for clustering technique used to determine the structure of the MLP neural network.
The comparative study presented in this section reinforce this method to support different types of training data such as interval, ordinal, categorical or mixture of different types of variables that ameliorate the ability to recognize clusters in the data. This improves the generalization capabilities of the proposed method.
Many datasets are analyzed where it was concluded that the experimental results obtained confirm those obtained by clustering program, which proves the validity of the proposed method.
Comparison of the proposed method with classical methods leads us to deduce that the proposed method performs well for the different type of datasets, which mean that is more flexible with different datasets types than classical methods. The proposed method adapts to the complexity of datasets to provide the best results regardless of the size of the dataset.
Using the proposed method, the design time and effort to determine the structure of the MLP neural network are reduced. This method can make the design simple and easy and possible for a Non-specialist designer.

4. Determination of optimal neural network Architecture using regression techniques

This chapter presents a new method to determine the architecture of a neural network based on regression methods. By clustering the training dataset, we obtain a set of parameters which can be useful to define regression models to determine the number of neurons in the hidden layers. The parameters obtained from clustering the training data can be taken as independent variables to determine the regression function. Data mining techniques represent a good tool to reveal and evaluate patterns in the training data to determine the complexity level of the problem to be solved by the neural network so that we can predict the optimal structure of a neural network. This method works unsupervised unlike other methods and more flexible with different datasets types. The proposed method adapts to the complexity of training datasets to provide the best results regardless of the size and type of dataset.
In order to the number of hidden units of a neural network, a Multiple Linear Regression models based on the parameters obtained from the clustering method described above is defined. In addition, a quality measure factor of the network architecture is defined based on the interconnection of layers is used.

4.1.The statistical hypothesis testing

The statistical hypothesis testing [16] are used to examine parameters obtained through hierarchical clustering of training dataset to select a number of parameters to determine the regression model. The probability coefficients of independent variables (P-value) have a value of less than 0.05 based on parameters proposed to be independent variables of the regression model. F-Test analysis used for the analysis of variance will be taken as evidence to prove that the structure of Multi-layer Perceptron depends on the selected factors.

The Multiple Determination Coefficient and the Multiple Correlation Coefficient results obtained [17] in both studies are near to one. It proves the validity and effectiveness of the defined models and the conciliator selection of factors included in the models.

Experimental data obtained using a clustering program are used to define regression functions to calculate the number of layers and the number of units in these layers of an MLP neural network. The parameters obtained using the clustering program are enough according to the statistical indicators to define the regression models and to guarantee the quality and accuracy of the defined regression models. The Multiple Linear Regression model has this form:

$$f = X \rightarrow Y$$
$$f(X) = w_0 + \sum_{j=1}^{n} w_j x_j \qquad (1)$$

The model defined to calculate the number of hidden neurons are based on the parameters which are the value of reference distance, the number of clusters obtained, the number of input forms of the MLP neural network, The number of grouped elements and the number of neurons in the input layer. Each regression model consists of 4 factors taken from the previous parameters.
Based on results obtained from algorithms used to determine the structure of neural network and based on information taken from the training examples it is seen that there is a mathematical relation between parameters used to determine the structure of MLP neural network and the results obtained with clustering algorithm. Moreover, the quality measure of the network structure was considered as an independent factor. The quality measure factor considers the

configuration and interconnection layers.
The proposed regression models consist of two models the first model used to determine the number of hidden layers and the second model used to determine the number of hidden units.

4.2.Stages of the Method Used

The proposed regression method depends on the results of pattern recognition algorithms applied on the data used to train the neural network seeks to define the number of hidden layers and the number of hidden units is evolved through several stages.
After applying the stages of the clustering method described in the previous chapter the next stage is setting up the number of hidden neurons. To determine the number of hidden layers and the number of neurons in the hidden layers, two Multiple Linear Regression models using parameters obtained in the previous steps has been developed. In addition, a quality measure factor of the network architecture that considers the configuration and interconnection layers defined in the previous step is introduced.
The comparison of the results obtained using the proposed regression method for the selected datasets with the results obtained by comparison of errors value for different neural network architectures to define the best architecture we confirm that the proposed regression method can predict the pest number of hidden layers and hidden neurons for a multilayer neural network.
The results obtained through the experimental tests confirm the validity of the proposed regression method to determine the neural network architecture based on the analysis of the selected dataset.
Conclusions:
It is noticeable that Pattern Recognition plays an important role in the determination of the optimal structure of the MLP neural network based on the proposed regression method.
By clustering the training dataset, we can collect a set of parameters useful to determine the structure of the MLP neural network as independent variables used to determine the regression models of the proposed method.
Using the proposed regression method to determine the structure of the neural network many factors affects the accuracy of the results. The main factor is Reference Distance which has a highest influence compared to the other factors. The value of Reference Distance must be selected precisely based on the criteria defined in the previous chapter.
We conclude that the proposed regression method to determine the structure of the MLP neural network in terms of the number of hidden layers and the number of neurons in these layers is viable. The model generated using the proposed regression method can be essential in practical applications and in the worst assumptions it can provide the initial number of hidden layers and neurons in these layers for an MLP neural network and relying on the information obtained from the training dataset the designer can reduce or increase these numbers.
This study proposes and develops a new design strategy for an MLP neural network which make the design of the structure of an MLP neural network unsupervised and helps to reduce the time allocated for network design.

5. Training Data Analysis to Improve Generalization Capabilities of Neural Network Architectures

This chapter presents an improvement for the generalization capabilities to the proposed clustering method presented in chapter 3 to determine the structure of a multi-layer neural network. The accuracy of the clustering method used requires precision upon the selection of clustering distance measures which necessitate a comparative study. The proposed clustering method depends on the results obtained through the clustering of the training data. Therefore, the analysis of training dataset is important to improve the generalization capabilities of Neural Network Architectures. Due to the diverse types of datasets, the classical methods do not cover all type of datasets. The proposed method is more flexible with various types of datasets in terms of the type of data, size, and number of features. Many aspects have been taken into consideration to give more generalization capabilities to the proposed method such as distance measures and linkage methods.

This chapter presents an amelioration for the generalization capabilities to the neural network architectures. By making a comparison of the results of diverse types of datasets, it was concluded that classical methods do not cover all type of training datasets. The effectiveness of the proposed method varies with the variation of the complexity of the problem that needs to be resolved. The proposed method can adapt to the complexity of the problem and have more flexibility with different types of datasets.

The proposed method imposes a perfect selection of distance measures due to the significant impact of the distance measures on the results of clustering of the training dataset, therefore the distance measures affect the accuracy of the architecture of Neural Network. Consequently, it is required a comparative study to determine the suitable distance measures. Distance metrics (such as Manhattan and Euclidean) will be accredited for calculation of the distance between observations. Linkage methods (such as average-linkage clustering, complete-linkage clustering, and single-linkage clustering) used to calculate the distance between objects will be applied with the Clustering method.

To prove the generalization capabilities of the proposed method we select a set of datasets with different numbers of instances and numbers of features. To cover different levels of complexity we compare the classification accuracy and error/epoch of the considered datasets to results obtained with different numbers of hidden layers until we can prove the effectiveness of the proposed method.

Based on the results obtained from various types of datasets with different numbers of instances and numbers of features. The number of hidden layers obtained using the proposed method get the lowest value of error/epoch for most of the datasets. The proposed method adapts to the complexity of dataset to provide the best results regardless of the size of the dataset. The results prove the generalization capabilities of neural network architectures defined using the proposed method.

Conclusions:

Usage of clustering techniques on the training data can define the optimal number of hidden layers. The precision is required when the selection of the appropriate clustering distance measures which necessitate a comparative study to perform a perfect clustering of training dataset where it affects positively on the accuracy of results.

The comparative study to determine the suitable distance measures gives to the proposed method more performance to handle different types of data which increase generalization capabilities of the neural network architectures.

The proposed method proves the capability to deal with various types of datasets, which enhance the generalization capabilities of the proposed method.

6. Comparison of the proposed method with classical methods

The design of the structure of an MLP neural network is a critical issue in the design of a Neural Network. Until now a generally valid theory to determine the structure of an MLP neural network based on the complexity of the problem is not specified. The methods actually used are evolutionary algorithms, exhaustive search, and Growing and Pruning algorithms, but they are very limited and time-consuming [18]. The more the number of hidden neurons was, the capacity of a neural network to solve problems increase but it can cause a long training time.

Decreasing the number of hidden neurons can cause an improves in the generalization of the neural network, but the neural network becomes weak or unable to meet the required needs.

In this chapter, several empirical formulas used to determine the number of neurons in the hidden layers of an MLP neural network will be examined and tested on a different training dataset. These classical methods will be compared to the proposed method as well. Several widely used formulas are selected for the comparison.

A set of datasets used such as Waveform Database Generator, Image Segmentation dataset, Glass identification, Landsat, Sonar, ECG, QRS, P-wave and T-wave datasets.

The proposed method gets the best percentage of accuracy for most datasets unlike the classical methods. Some classical formulas perform well with small datasets which have a few training items. Whereas other classical formulas perform well with large data sets because they take into consideration the number of training items. The formulas which depend mainly on the number of input and output neurons are effective for small datasets while do not perform well with large datasets which have complex problems to solve. The results of error/epoch obtained is almost similar to the result of percentage of accuracy.

Comparison of the proposed method with classical methods lead us to deduce that the proposed method performs well for different type of datasets, which mean that the proposed method is more flexible with different datasets types than classical methods. The proposed method adapts to the complexity of dataset to provide the best results regardless the size of dataset. In some cases, the dataset is chosen with a size more than required, which leads to a bad results using classical methods but this problem is avoided by using the proposed method since it focus on the complexity of problem to be solved regardless the size of dataset.

7. Conclusion and Personal Contributions

Neural networks represent an important tool in data classification. It is the only technique that allows generalizations based on a set of data to be analyzed. Regardless of the chosen neural network architecture and how learning is used, the number of hidden layers and the number of hidden units are currently determined empirically by testing.
In this thesis, we demonstrated the importance of data mining techniques to determine the architecture of the neural network. We believe that the analysis of the dataset used to train the neural network can lead to the optimal neural network architecture. Patterns discovered in the training dataset can determine the complexity level of the considered problem through it we can define the architecture of the network. In this thesis, we exploit the Data Mining techniques to analyze the training dataset to determine a general method, which adapts with all type of problems considered by the neural network.

The following objectives have been achieved:

1. An analysis of the currently used methods to determine the architecture of neural network was carried out. Several widely used methods to determine the architecture of the neural network was discussed and the weaknesses of these methods are presented.
2. Determining the extent to which the structure of a neural network affect the learning ability of the network. This analysis sought to define the relationship between the

number of hidden layers/neurons and the complexity of the problem to be solved by the neural network. The analysis has also expanded on how the architecture of neural network must adapt to improve the generalization capabilities of the network.

3. The study was carried out, whose purpose was to find a way to exploit Data mining techniques to analyze the dataset used to train the neural network to extract useful information, and establish relationships based on pattern discovered from data to find solutions for determining to optimal neural network architecture.
4. The study of Data mining techniques was carried out, whose purpose was to cluster the dataset used to train a neural network based on a set of pattern recognition techniques to extract common features and to correlate the number of groups obtained with the optimal number of hidden layers for a multilayer neural network. Several criteria were defined according to the context in which to correlate the number of the obtained cluster with the optimal number of hidden layers.
5. After determining the number of hidden layers based on the proposed clustering method, to determine the number of hidden neurons also become necessary. The study purpose was to select a set of parameters obtained using the proposed clustering method which can be useful to define regression models to determine the number of hidden neurons. Depending on the statistical hypothesis testing and based on the F-Test analysis, it has been proven that the number of hidden neurons is dependent on the considered factors.

The research results, which was carried out over several stages, was finalized by proposing a new general method of to determine the optimal neural network architecture using Agglomeration Hierarchical Clustering Algorithm and Regression methods.

Use of the Agglomeration Hierarchical Clustering Algorithm can help extract several common characteristics of input forms by which neural network input records are classified. Using the clustering method, a number close to the optimal number of hidden layers of a multi-layer neural network can be determined.

This study proposes and develops a new design strategy for an MLP neural network where make the design of the structure of an MLP neural network unsupervised and helps to reducing the time allocated for network design.

Using the proposed method, the design time and effort to determine the structure of the MLP neural network are reduced. This method can make the design simple and easy and possible for a Non-specialist designer.

The optimal architecture of a neural network can be determined using the results of the clustering algorithm applied to the training data. Using the proposed clustering method, many factors affect the number of hidden layers of a multi-layer neural network. One such factor is the reference distance. The accuracy of this factor is important in determining the optimal number of hidden layers.

Using the proposed regression method to determine the structure of the neural network many factors affect the accuracy of the results. The main factor is Reference Distance which has an important effect compared to the other factors. The value of Reference Distance must be selected precisely based on the criteria defined in the previous chapter.

In order to show this method's viability for representation, a study in Chapter 3 was conducted in the thesis. Viability consisted of a comparative study of clustering distance measures lends on the proposed method more effectiveness and accuracy through the perfect selection of the appropriate distance measures for clustering technique used to determine the structure of the MLP neural network.

It is evident from this research that Pattern Recognition has a considerable role in the

determination of the structure of a neural network.
In the study conducted in Chapter 4, experiments of the proposed regression method were performed according to the statistical indicators, which proven that the model generated using the proposed regression method is viable and it can be essential in practical applications.

The following conclusions resulted:

- The statistical hypothesis testing proved that the regression model is significant. Hypothesis testing demonstrated that the independent variables obtained through hierarchical clustering of training dataset have a significant relationship with the dependent variable. Therefore, a relationship between the clustering results and the structure of Multi-layer Perceptron exists. Consequently, the regression model was established.
- Through this study, we conclude that the proposed regression method to determine the structure of the MLP neural network in terms of a number of hidden layers and the number of units in these layers is viable.
- The model generated using the proposed regression method can be essential in practical applications and at least it can provide the initial number of hidden layers and neurons in these layers for an MLP neural network and relying on the information obtained from the training dataset the designer can reduce or increase these numbers.
- The proposed clustering method can define the structure of the neural network with good accuracy results. To improve generalization capabilities of the method it required a precise selection of the appropriate clustering distance measures. Therefore, a comparative study of the distance measures used in clustering of the training data is necessary to perform a perfect clustering, which affects positively on the accuracy of results of the proposed method.
- To increase generalization capabilities of the proposed method we need a comparative study to determine the suitable distance measures in order to enhance the performance to handle different types of training data.
- The proposed method proved the capability to deal with various types of training datasets, which enhance the generalization capabilities of the proposed method.
- Comparison of the proposed method with classical methods leads us to deduce that the proposed method performs well for the different type of datasets, which mean that the proposed method is more flexible with different datasets types than classical methods. The proposed method adapts to the complexity of training datasets to provide the best results regardless of the size of the dataset.

The contributions for this thesis refer to:

1. Theoretically defining a new, original method for determining the optimal neural network architecture through the analysis of the training dataset.
2. One of the most important contributions of this thesis is developing an algorithm to determine the architecture of the neural network based on the complexity of the problem to be solved.
3. In this study, a third-party platform of Weka based on the proposed method was developed and used.

4. The proposed method adapted to the complexity of the training data to provide the best results regardless the size and type of dataset in contrast to the currently used methods.
5. A comparative study of the proposed method with the currently used methods leads us to deduce that the proposed method performs well for the different type of datasets, which mean that the proposed method is more flexible with different datasets types than other methods.
6. The comparative study of clustering distance measures lends on the proposed method more effectiveness and accuracy through the perfect selection of the appropriate distance measures for clustering technique used to determine the structure of the neural network.

The results obtained were validated by publishing 5 articles, of which 4 are ISI indexed in Web of Science, and three indexed IEEEs. The published articles have been cited 8 times.
In the present research, a series of papers have been presented and published which follow and reflect the various stages of research that have been undertaken. These works are:

1) Tej, Mohamed Lafif, and Stefan Holban. "Determining Multi-layer Perceptron Structure Using Clustering Techniques." International Journal of Artificial Intelligence 17, no. 1 (2019): pp. 139-166.

2) Tej, Mohamed Lafif, and Stefan Holban. "Determining neural network architecture using data mining techniques." In 2018 International Conference on Development and Application Systems (DAS), pp. 156-163. IEEE, 2018.

3) Tej, Mohamed Lafif, and Stefan Holban. "Comparative Study of Clustering Distance Measures to Determine Neural Network Architectures." In 2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 000189-000194. IEEE, 2018.

4) Tej, Mohamed Lafif, and Stefan Holban. "Determining optimal neural network architecture using regression methods." In 2018 International Conference on Development and Application Systems (DAS), pp. 180-189. IEEE, 2018.

5) Tej, Mohamed Lafif, and Stefan Holban. "Determining Optimal Multi-Layer Perceptron Structure Using Linear Regression." In International Conference on Business Information Systems, pp. 232-246. Springer, Cham, 2019.

Bibliography

[1] A. Itamar, D. C. Rose and T. P. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research," *IEEE computational intelligence magazine*, vol. 5, no. 4, pp. 13-18, 2010.

[2] Y. Zhao, J. Chen and H. Vincent Poor, "Efficient neural network architecture for topology identification in smart grid," in *IEEE Global Conference on Signal and*

Information Processing (GlobalSIP), 2016.

[3] C. Bishop, Pattern recognition and machine learning, Springer, 2006.

[4] J. Bongard, "Biologically Inspired Computing," *INSPEC Accession Number: 10712660,* vol. 42, no. 4, 2009.

[5] H. Guanshan, "Neural Network Applications in Sensor Fusion for a Mobile Robot Motion," in *WASE International Conference on Information Engineering*, 2010.

[6] R. Lovassy, L. T. Kóczy and L. Gál, "Fuzzy neural networks stability in terms of the number of hidden layers," in *IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2011.

[7] K. Shibata and Y. Ikeda, "Effect of number of hidden neurons on learning in large-scale layered neural networks," in *ICCAS-SICE*, 2009 .

[8] M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, second ed., 2005.

[9] Z. Haoyang, M. D. Edwards, G. Liu and D. K. Gifford, "Convolutional neural network architectures for predicting DNA–protein binding," *Bioinformatics,* vol. 32, no. 12, pp. i121-i127, 2016.

[10] P. B. Z. a. Q. V. L. Ramachandran, "Searching for activation functions," *arXiv preprint arXiv:1710.05941,* 2017.

[11] Z. Chiyuan, S. Bengio, M. Hardt, B. Recht and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530,* 2016.

[12] D. Ian and S. S. Ravi, "Agglomerative hierarchical clustering with constraints: Theoretical and empirical results," in *European Conference on Principles of Data Mining and Knowledge Discovery, pp. 59-70. Springer*, Berlin, Heidelberg, 2005.

[13] B. Pavel., "A survey of clustering data mining techniques," in *Grouping multidimensional data, pp. 25-71. Springer*, Berlin, Heidelberg, 2006.

[14] D. Gopikrishna, P. Wang, D. Rangaprakash and B. Wilamowski, "Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data," *IEEE transactions on cybernetics* , vol. 45, no. 12, pp. 2668-2679, 2015.

[15] S. A. a. a. T. Y. W. A. S. Shirkhorshidi, "A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data," *PLOS ONE,* vol. 10, no. 12, 2015.

[16] P. Francesc and Y. Vidal., "Wind turbine fault detection through principal component analysis and statistical hypothesis testing," *Energies,* vol. 9, no. 1, p. 3, 2016.

[17] S. Chatterjee and A. S. Hadi., Regression analysis by example, John Wiley & Sons, 2015.

[18] T. A. J, P. M, D. W. S, M. G. S and E. M. R, "On Predicting the Optimal Number of Hidden Nodes," in *,” 2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2015.