NEURAL NETWORK ARCHITECTURE THROUGH DATA MINING TECHNIQUES

MOHAMED LAFIF TEJ

Politehnica University of Timişoara Romania
E-mail: afiftej@gmail.com

Abstract— The paper presents the technique used to determine the optimal neural network architecture by means of Pattern Recognition and Data Mining Techniques. Using clustering techniques on the data to be processed 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.

Index Terms—Clustering methods, data mining, hidden layers, neural network architecture, pattern recognition.

I. INTRODUCTION

The issue that I want to address is a complex issue and difficult because there are no generally available analytical methods for determining the optimal number of hidden layers to a multilayer neural network [1]. Among the methods currently used to determine the optimal neural network architecture as choosing multiple networks with different architectures and their learning on a common data set [2] [3], until they meet the performance and then, choosing that particular neural network with minimum size as the number of layers and neurons can possibly take into account the training time [4]. The disadvantage of the method is that the results depend largely on the structures originally selected and time consuming. In this paper we propose an original method for determining the architecture of an optimum neural network (number of hidden layers) using clustering techniques that are part of the broader field of artificial intelligence, namely in the Pattern recognition.

II. CLUSTERING TECHNIQUE USED

Clustering algorithm used is Kruskal's algorithm. This grouping based on data analyzed tree with minimum distance. A tree with minimum distance is a tree on the set E, which $\sum d(i, j)$ is minimum. After its construction, clusters obtained by cutting all segments whose length is greater than a reference value chosen by the user called Reference Distance. The Figure 1 illustrates the grouping technique using Kruskal's algorithms [5].

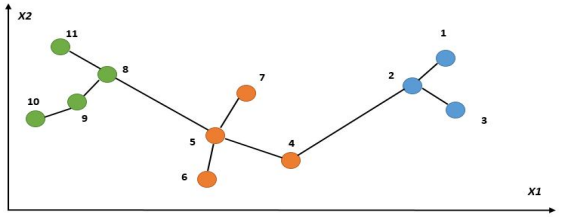


Fig.1. Clustering technique that uses Kruskal's algorithm using tree with the minimum distance, it is observed that the obtained three groups of elements (1, 2, 3), (4, 5, 6, 7) and (8, 9, 10, 11) according to the chosen reference distance.

III. DETERMINING OF THE NUMBER OF HIDDEN LAYERS USING CLUSTERING TECHNIQUES

The idea behind this method of determining the optimal partition starting from the set of N training forms for a multi-layer neural network. We assume that the set of N forms used to train a neural network can be grouped together using conventional methods of pattern recognition [6] [7] according to a specific criterion so obtaining a number of homogeneous classes and a number of items that could not be associated with any group elements unclassified dispersed. The number of hidden layers of a multi-layer neural network will equal the number of groups obtained by grouping when the system is stable. Therefore, the number of classes obtained when the process is stable, will be equal to the number of hidden layers of a multilayer neural network [8] [9]. To determine the number of hidden layers we use clustering technique described above.
Figure 2 presents the number of groups obtained for Landsat Satellite images dataset [10] for a reference distance of 75.

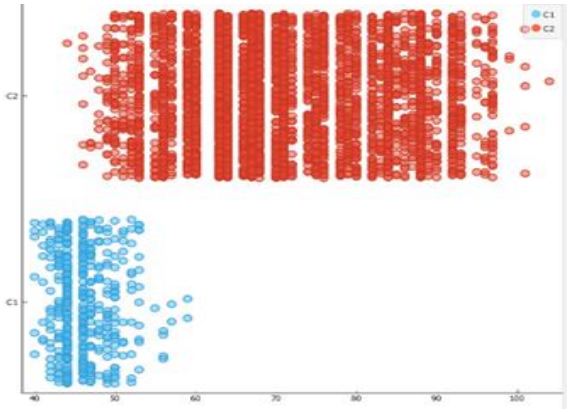


Fig.2. The number of groups and the number of dispersed elements obtained the Landsat Satellite images for a distance of the reference d = 75, 2 groups obtained.

In Table 1 given results with the program of grouping to classify Landsat Satellite images of the same scene in different spectral bands.

TABLE 1: DATA OBTAINED WITH GROUPING PROGRAM FOR LANDSAT SATELLITE IMAGES DATA SET

Reference Distance	Number of clusters	Number of items grouped in%
75	2	90%
67	2	80%
58	2	70%
50	3	59%
35	4	41%
32	5	38%
30	6	36%

Figure 2 shows the connection between the reference distances for which the group and the number of groups obtained for the Landsat Satellite images, including the percentage of forms that are presented and clustered.

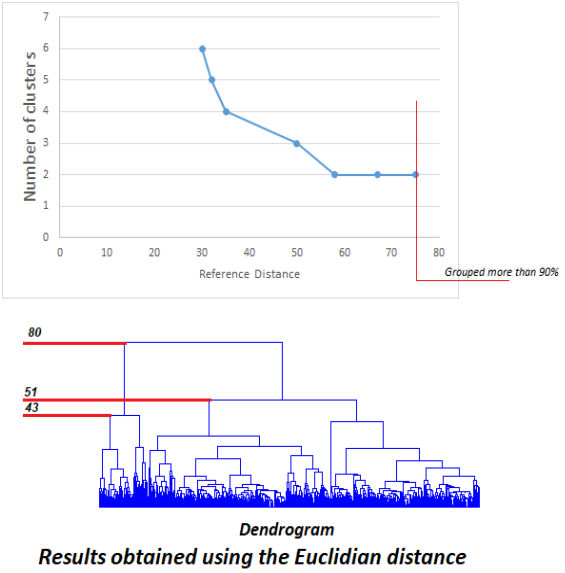


Fig.3. The relations between the number of groups that obtained according to the reference distance (Euclidean) and corresponding Dendrogram obtained from the analysis of Landsat Satellite images

From the above it is noted that with increasing distance of reference obtained increasingly fewer groups and a higher percentage of clustered elements. For a large enough distance of the reference all forms of input are clustered into one group.

The idea is to stop at the reference from which we get a grouping of as many forms (as few forms that cannot be clustered), namely more than 90%. The number of clusters so obtained we are going to consider the optimal number of hidden layers for the neural network.

Therefore, in the table above we can conclude that the analysis of Landsat Satellite images dataset would have a neural network with two hidden layers.

It easily seen in the importance of each parameter in determining the number of hidden neurons for a multilayer neural network.

The number of neurons in the hidden layers of a multilayer neural network influenced by several factors, the most important being the number of hidden layers and the number of inputs in multilayer neural network [11].

IV. EXPERIMENTAL RESULTS

For the analysis of the Datasets used by a neural network until we can define the optimal multi-layer neural network architecture, we test each data set. These data tested with different multilayer neural networks architectures using supervised learning method, in particular backpropagation learning algorithm.

Simulation and training neural networks was done using simulation program "Neuroshell ver. 2.0" of Ward Systems Group company program [12] which considered one of the most complete and advanced simulators in neural networks.

For data analysis, we used three different types of neural network architectures, with one hidden layer, two hidden layers and three hidden layers.

Figure 4 shows the structure of three types of neural network architectures used.

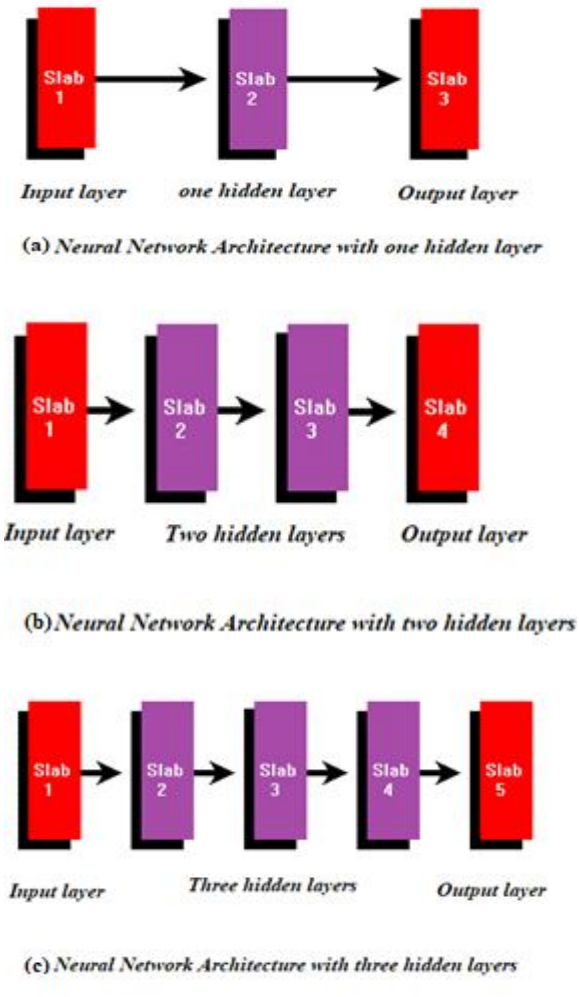
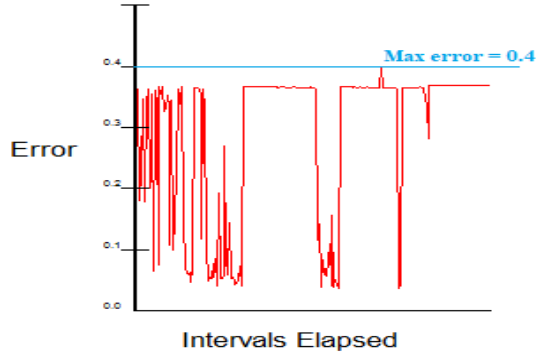
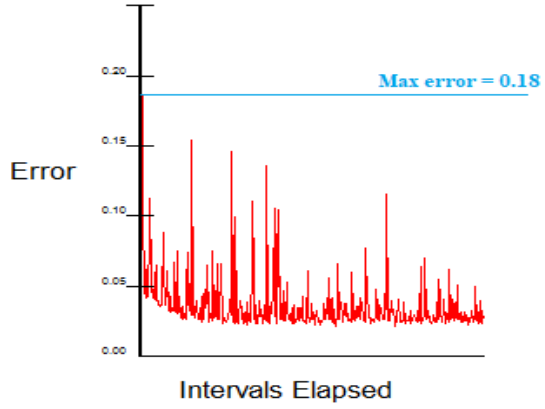


Fig.4. the three types of neural network architectures used for data analysis

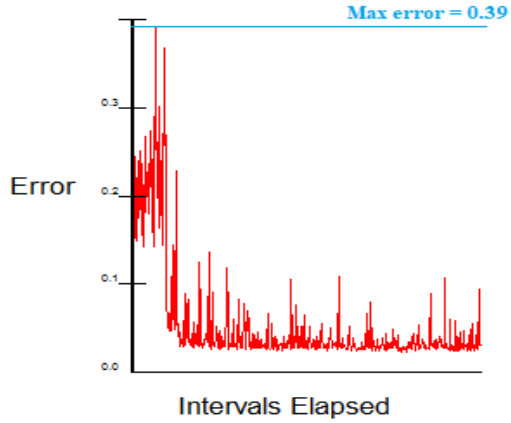
We will take Landsat Satellite images dataset as an example to demonstrate the differences between the error results for each neural network architecture where the error is the difference between actual and desired output values.
The figure 5 are presenting a comparison of Landsat Satellite errors obtained by different neural network architectures.



(i) The average error for Neural Network Architecture with one hidden layer



(ii) The average error for Neural Network Architecture with two hidden layers



(iii) The average error for Neural Network Architecture with three hidden layers

Fig.5. the relative error for the Landsat Satellite obtained by neural network architectures with different numbers of hidden layers

Please note that for results of Network with one hidden layer has the largest number of errors with value more than 0.35 and for the network with two hidden layers has one answer with a larger error less than 0.18. For network with three hidden layers we have two errors greater than 0.30.
The experimental results for Landsat Satellite images make us conclude through error comparison obtained by different neural network architecture, that network with two hidden layers have an error less than networks with one or three hidden layers.
From the results of the analysis of Landsat Satellite images, we notice that the best performs is for a neural network with two hidden layers. The results obtained confirm those obtained by clustering program that presented previously.
Table 3 below show a comparison of the results in terms of error for six datasets [13] [14] with those obtained by clustering techniques that presented previously.

TABLE 3: COMPARISON BY BEST NUMBER OF HIDDEN LAYERS OBTAINED BY DIFFERENT NEURAL NETWORK ARCHITECTURES AND THE NUMBER OF CLUSTERS OBTAINED BY CLUSTERING TECHNIQUES FOR ALL DATASETS.

Data Set	Number of clusters	Best number of layers obtained by different neural network architectures
Landsat Satellite	2	2
P-wave	2	2
QRS	1	1
T-wave	3	3
ECG	1	1
Sonar	2	2

Based on the analysis of experimental results obtained in this section we confirm the optimal architecture of a multilayer neural network can be determined using the clustering program for the six sets of data analyzed as it presented previously.

CONCLUSION

Neural Networks represent an important tool in data classification, is the only technique that allows generalizations based on a set of data to analyze. Regardless of chosen neural network architecture and how learning used, determination of numbers of hidden layers and the number of neurons in these layers is empirically by testing.
Clustering techniques highlighted a number of common input forms characteristics, on which neural network input forms classified into groups (clusters).
We analyze the data processed by a multilayer neural network 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.
Using clustering techniques many factors effects on the number of hidden layers of a multilayer neural network, the most important is reference distance according it input data grouped. The accuracy of this

factor is important to determine the optimal number of hidden layers.

We believe that using this method can save time in building an optimal neural network for solving a concrete problem.

REFERENCES

[1] Michael Negnevitsky. "Artificial Intelligence - A Guide to Intelligent Systems". Addison-Wesley, second edition, 2005.

[2] "Irvine Machine Learning Dataset and Benchmark," Center for Machine Learning and Intelligent Systems, [Online]. Available: http://archive.ics.uci.edu/ml/datasets.html. [Accessed May 2016].

[3] Ian. T. Nabney, Algorithms for Pattern Recognition, London: Springer, 2004.

[4] Sulee Bunjongjit, Atthapol Ngaopitakkul, Chaichan Pothisarn and Chaiyan Jettanasen, "Improvement to reduce training time of back-propagation neural networks for discrimination between external short circuit and internal winding fault," in International Conference on Information Science, Electronics and Electrical Engineering, 2014.

[5] Raoul Berger, Séverine Dubuisson and Christophe Gonzales, "Fast multiple histogram computation using Kruskal's algorithm," in 19th IEEE International Conference on Image Processing, 2012.

[6] Christopher Bishop, Pattern recognition and machine learning, Springer, 2006.

[7] Hongfei Pan, Dong Liang, Jun Tang, Nian Wang, Wei Li, "Shape recognition and retrieval based on edit distance and dynamic programming" Tsinghua Science and Technology (Volume: 14, Issue: 6, Dec. 2009).

[8] Carely Guada, Daniel Gómez and J. Tinguaro Rodríguez, "Fuzzy Image Segmentation Based on the Hierarchical Divide and Link Clustering Algorithm," in 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2015.

[9] Grady Leo, and Eric L Schwartz, "Isoperimetric Graph Partitioning for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 3, p. 469–475, 2006.

[10] Statlog (Landsat Satellite) Data Set, UC Irvine Machine Learning Repository.

[11] Takahiro. Sasaki, Kentaro Kinoshita and Satoru Kishida, "Effect of number of input layer units on performance of neural network systems for detection of abnormal areas from X-ray images of chest," in IEEE 5th International Conference on Cybernetics and Intelligent Systems (CIS), 2011.

[12] "NeuroShell 2," Ward Systems Group, Inc, [Online]. Available: http://www.wardsystems.com/neuroshell2.asp.J

[13] "Arrhythmia Data Set," UC Irvine Machine Learning Repository.

[14] "Connectionist Bench (Sonar, Mines vs. Rocks) Data Set," UC Irvine Machine Learning Repository.

★★★