Determining Optimal Neural Network Architecture Using Regression Methods

Mohamed Lafif Tej, Stefan Holban
Politehnica University of Timisoara, 300223 Timisoara, Romania
afiftej@gmail.com, stefan.holban@cs.upt.ro

***Abstract*—This paper presents regression methods used to determine the optimal neural network architecture. Using the results obtained from clustering the input data as independent variables to define the regression function. This function will determine the optimal number of hidden layers and the optimal number of neurons on each hidden layer for a multi-layer neural network. Data mining techniques are used to explore input 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.**

Keywords—Clustering Methods, Data Mining, Multi-layer neural network, Neural networks, Pattern Recognition, Regression analysis

I. INTRODUCTION

Regression is a data mining function which is used to examine the relationship between one dependent and one or more independent variables. Regression techniques are used to determine variable changeable dependencies when analytical methods fail to do so.

The basic form of regression models includes unknown parameter (β), independent variable (X), and the dependent variable (Y). The regression function $Y \approx f(X, \beta)$ is used to predict the values of (Y), if the value of (X) is given.

Regression analysis is used for predicting and forecasting. It is frequently used in the field of machine learning. Regression analysis is used to infer causal relationships between one dependent and one or more independent variables under restricted circumstances. Neural Network [1-5] can use regression techniques as a solution to determine the optimal architecture of a multi-layer neural network.

A common criticism of Neural Networks is there are no generally available analytical methods for determining the optimal number of hidden layers and the number of neurons in these layers for a multi-layer neural network [6-11]. There are several algorithms for determining architecture of neural networks, but only few of them can successfully give accurate architectures under specified conditions [7], [12-13].

In this paper, an original method will be proposed to determine the number of hidden layers and the number of neurons in each hidden layer for a neural network using data mining techniques represented by regression and clustering techniques [14-15] depending on pattern recognition.

This paper is organized as follows. Section 2 presents Method used to determine the Architecture of a Neural Network. Section 3 describes the definitions of data mining techniques used. Section 4 discusses clustering techniques used to determine the architecture of a neural network. Section 5 discusses the regression method used to improve predictability of the optimal neural network architecture based on results obtained from clustering techniques. Section 6 presents a comparison between experimental results and discussion of these results. Section 7 concludes this work.

II. THE METHOD USED TO DETERMINE THE ARCHITECTURE OF A NEURAL NETWORK

A. Stages of the Method Used

This method, which is based on pattern recognition techniques, presents how to determine the number of hidden layers and the number of neurons in these layers. The algorithm for dimensioning the neural network is evolved through a number of stages:

1) Establishing the training set: Establishing the training set upon which the neural network learning will be done. Initially, it is pre-processed to eliminate noise, incomplete records and those records showing large dissimilarities to other data.

2) Establishing the number of input neurons: The number of input layer neurons will be equal to the number of features [30].

3) Determining the number of output neurons: The output layer will have a number of neurons as the number of desired classes in the case of classification problems, or the number of output features vector in other cases.

4) Determining the number of hidden layers: The number of hidden layers is equal to the number of clusters in which the analyzed input forms can be classified when these two required conditions are met:

- Grouping of at least 90% of the items in the dataset used to train the neural network.
- The process of grouping is stable.

In this study, the experiment has shown that to determine the optimal neural network architecture, the number of hidden layers will be equal to the number of obtained clusters when the above two conditions are met.

978-1-5386-1495-2/18/$31.00 ©2018 IEEE

5) Determining the number of neurons in hidden layers: In order to calculate the number of neurons in the hidden layers, a multiline regression model using parameters obtained in the previous step has been developed. In addition, a quality measure factor of the network architecture that takes into account the configuration and interconnection layers defined in the previous step are introduced.

Regression analysis is a statistical technique identifying the most appropriate equation of the regression function that expresses a quantitative relationship between a dependent variable and one or more independent variables. The method requires the existence of a base of experimental data obtained by observations or measurements. In order to characterize the intensity of the relationship between the variables y and x, the indicator of multiple determination coefficient (ratio) R^2 is required, which expresses which expresses the level of variation of y affected by the variation of x. The indicator values of Multiple Determination Coefficient R^2 is between 0 and 1. If the value is close to 1, the variation of the variable y affected by the variable x is high and therefore the intensity of the relationship between the two variables is high as well. This indicator is important but it must be correlated with other statistical indicators for a better assessment of the regression model.

To determine the regression function by which the number of hidden layers and the number of hidden neurons are determined for multi-layer neural networks, experimental data obtained using the clustering algorithm are used. Data obtained with this algorithm are sufficient according to statistical indicators calculated for building the regression model and to ensure the quality and reliability of the built model. To build the mathematical model, a Multiline Regression Model is chosen having this form:

$$y = a + a_1x_1 + \ldots + a_ix_i + \ldots + a_nx_n \quad (1)$$

The constructed mathematical model for determining the number of hidden layers has four factors, the most important being the Reference Distance for which the grouping is made. To determine the number of neurons on the hidden layers, the model has the four most important parameters which are the Reference Distance for which the grouping is made, the number of hidden layers and the number of inputs of the multi-layer neural network.

The following conclusions are drawn from the two studies:

- Depending on the statistical hypothesis testing [31], the probability coefficients of independent variables that occur in chosen models is below 5%. Therefore, these independent variables cannot be eliminated from the model.
- Based on the F-Test analysis (used for the analysis of variance), it is concluded that the number of hidden layers and the number of neurons in these hidden layers are dependent on the factors taken into account.
- There are links between all factors in each model. Moreover, a relatively small positive and negative correlations exist, but they are insignificant.
- The values obtained for the Multiple Determination Coefficient and the Multiple Correlation Coefficient in both studies are close to 1. It reflects the effectiveness of the proposed model and the right choice of factors included in the model.

The dimensioning algorithm of a neural network starting from training examples shows that there is a mathematical relation between parameters that define the architecture of a multi-layer neural network (number of hidden layers and the number of neurons in these layers) and the results obtained with the clustering algorithm:

- The number of clusters
- The number of grouped elements
- Reference Distance
- The number of training forms
- The number of neurons in the input layer

B. Materials Used

To determine the regression function, DataFit program is used. This program allows to determine a regression function with up to 20 independent variables and generate a large variety of functions that covers most of the regression models used. DataFit program is not only used to determine the expression of regression functions, but also allows to draw conclusions about the effectiveness with which the independent variables are chosen and the quality and credibility of the regression model selected.

The results are validated by analyzing experimental data obtained by the statistical program SPSS version 15.0, which have given identical results.

III. DATA MINING TECHNIQUES USED

Data Mining is an analytic process designed to explore large datasets to identify patterns and establish relationships to solve problems through data analysis. The main factor of data mining is prediction represented by the ability to reveal hidden patterns and relationships in data that can lead to reasonable predictions.

In this paper, there is a number of Data Mining branches used on the input data to determine the optimal architecture of a neural network, including Anomalies Detection, Clustering, Classification and Regression.

A. Clustering Algorithm Used

The data to be processed by the neural network can be clustered to extract common features by means of pattern recognition techniques [16-17].

The purpose of this paper is to determine the architecture of a neural network depending on the results obtained by applying Agglomerative Hierarchical Clustering algorithm on the data that will be processed by the neural network. Cluster Analysis, based on the algorithm of Agglomerating Hierarchical Clustering, partitions information into groups. Grouping is the

process of grouping objects based only on the information describing the objects and the relationships between them.

The goal behind grouping is to improve level of similarity of objects inside the group and decrease as much as possible similarity of objects in different groups. If a group contains several subgroups then there will be a hierarchical grouping. This hierarchical grouping is a set of nested groups that are organized in a tree-like form. Each tree node (group) represents a group of its subgroups. The root of the tree is the group that consists of all other groups and objects. Tree leaves may in some cases represent groups of one object. The hierarchical grouping can be seen as a sequence of groups that can be partitioned. The partitioning process can be achieved by removing members from a sequence.

The Agglomerative Hierarchical Clustering algorithm [18-22] is a collection of closely related clustering techniques that perform a hierarchical grouping by starting from each object as a single group, and then by repeatedly joining the closest groups until one remains. There are two basic methods to generate a hierarchical grouping, which are Agglomerative and Divisive methods. The Agglomerative Hierarchical Clustering is the most used grouping technique. This type of grouping can be displayed using a Dendrogram that will display different grouping details. One such example is represented in Fig. 1 [23].

The number of clusters are obtained by cutting all the segments whose length is greater than a reference value chosen by the user. This number will be used to determine the architecture of a neural network. In this paper, the reference value will be referred to by "Reference Distance".

Fig. 1 presents segments cut from a Dendrogram for defined values of Reference Distance.

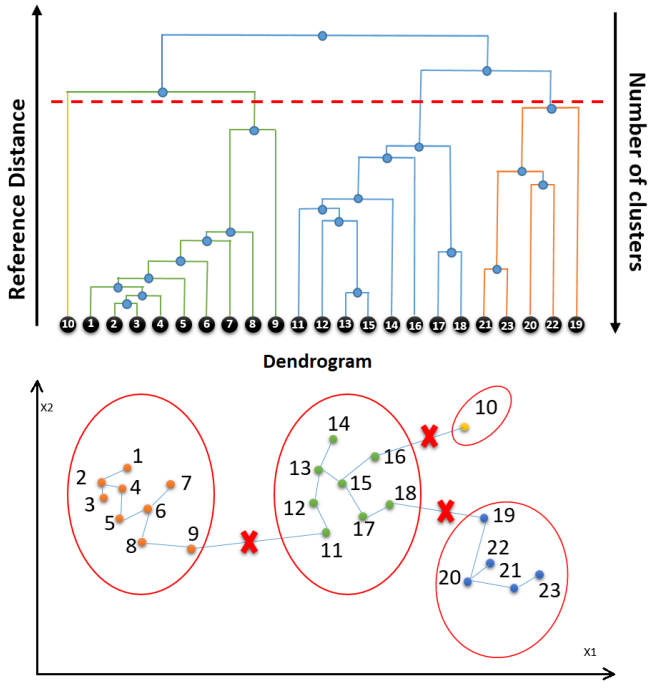


Fig. 1. obtained groups based on the chosen value of Reference Distance using Agglomerative Hierarchical Clustering algorithm..

B. *Regression Method Used*

Another data mining technique used in this paper to determine the architecture of a neuronal network by means of the regression methods [24-25]. The experimental results obtained by grouping the input data are considered useful in accordance with the purpose to determine the number of hidden layers and the number of neurons on each layer for a multi-layer neural network.

To give the predicted number of hidden layers obtained by clustering techniques more accuracy and appropriateness with the possible changes and anomalies in the patterns found in input data, the use of the regression method is convenient.

To determine a regression function, a set of experimental data needs to be obtained by observations or measurements that will show how the variation of the dependent variable y depends on the changes in independent variables x1, x2 …xn.

Regression methods can determine the architecture of a multi-layer neural network using the result obtained from clustering input data and applying pattern recognition techniques.

IV. Clustering Techniques Used to Determine the Architecture of Multi-Layers Neural Network

In this method, which is based on pattern recognition techniques, presents how to determine the number of hidden layers and the number of neurons in these layers. The idea is to group the training forms (input data) of a neural network using conventional methods of pattern recognition [26-27] according to a specific criterion to obtain a number of groups.

Using information generated by clustering techniques have shown that the number of groups obtained when at least 90% of the forms of input are grouped is equal to the number of hidden layers required to obtain an optimal neural network architecture.

Number of groups should be taken as few as possible until a network is retrieved with fewer hidden layers to reduce the complexity of a neural network architecture. The optimal number of hidden layers will be equal to the number of groups obtained through clustering techniques. Therefore, the number of classes obtained from input forms will be equal to the optimal number of hidden layers.

The process of grouping must be stable, which means by increasing the Reference Distance the number of groups does not change with the proviso that the extreme cases not taken into account (a very short distance for which each element represents a group or relatively large distances to all the elements placed in a one group).

Using clustering techniques on Forest Type Mapping dataset [28-29] to determine the number of clusters, which will be taken as the optimal number of hidden layers of a multi-layer neural network for this dataset. The Forest Type Mapping dataset contains multi-temporal remote sensing data of a forested area in Japan. The goal is to map different forest types using spectral data. It is composed of 326 instances. The Forest Type Mapping dataset contains training and testing data from a remote sensing study, which mapped different forest types

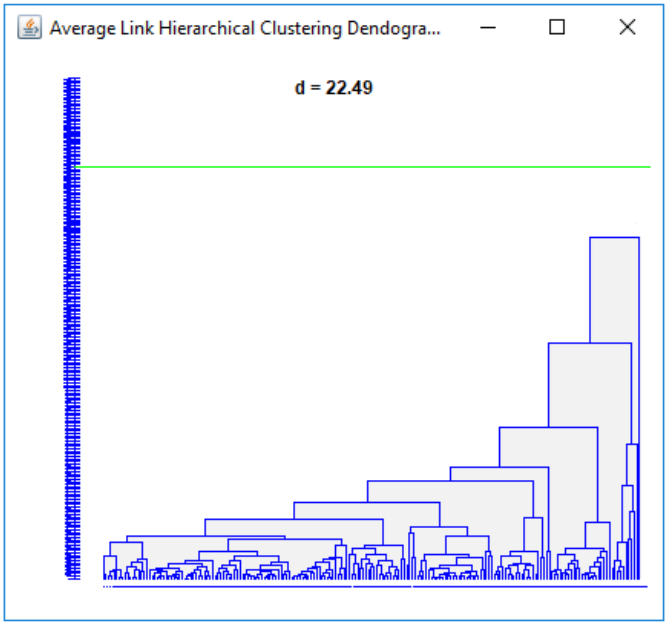
based on their spectral characteristics at visible-to-near infrared wavelengths, using ASTER satellite imagery. The output (forest type map) can be used to identify and/or quantify the ecosystem services (e.g. carbon storage, erosion protection) provided by the forest.

Table 1 presents the number of groups and percentage of elements clustered for different values of the Reference Distance.

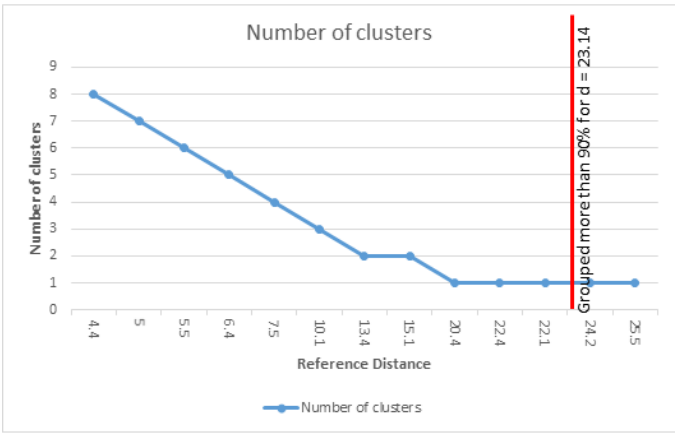
TABLE I. NUMBER OF GROUPS AND PERCENTAGE OF ELEMENTS CLUSTERED FOR DIFFERENT VALUES OF REFERENCE DISTANCE

Reference Distance	Number of clusters	Number of items grouped in%
9.415	1	97.2
8.885	1	91.77
7.698	1	79.51
6.355	1	65.64
5.34	2	55.15
3.279	3	33.87
1.671	4	17.26
1.499	5	15.48
1.437	6	14.84

Fig. 2 shows the number of groups obtained based on the Reference Distance used and Dendrogram associated with clustering process for classification of the Forest Type Mapping dataset.



(i) Dendrogram



(ii) Number of groups obtained based on the reference distance values

Fig. 2. Number of groups obtained based on the Reference Distance used and Dendrogram associated with clustering of the Forest Type Mapping dataset.

From the studies described above and analysis of the results obtained by clustering of the Forest Type Mapping dataset it can be concluded that a high value of Reference Distance has resulted in a low number of groups and a large number of clustered elements, and for the largest value of Reference Distance all elements are clustered into one group.

A value of Reference Distance is set when the grouping covers at least 90% of the analyzed forms. From the analysis results shown in Table 1 it can be seen that the best value of Reference Distance is 8.885 corresponding to a number of one cluster. This means the analyzed forms contain one group based on studies described above. The number of groups obtained will be taken as the number of hidden layers. Therefore, we can conclude that a neural network with one hidden layer is optimal for the analysis of the Forest Type Mapping dataset.

V. REGRESSION TECHNIQUES USED TO DETERMINE THE ARCHITECTURE OF MULTI-LAYERS NEURAL NETWORK

In this step, to improve predictability of results obtained with the proposed clustering methods, a multiline regression model will be constructed.

In order to calculate the number of hidden layers and the number of hidden neurons a multiline regression model using parameters obtained from the previous step has been developed. In addition, a quality measure factor of the network architecture that takes into account the configuration and interconnection layers has been introduced.

A. Factors that determine the number of hidden layers of a multi-layer neural network

A set of experimental data are required to determine the regression function. This set of data presents the variation of dependent variable y (number of hidden layers) depending on changes in independent variables x1, x2 …xn.

The training set of a neural network represented by a set of examples consisting of an input vector and possibly an output vector on which the neural network can learn to solve some

types of problems. The experimental results obtained previously are found to be useful in accordance with the purpose of determining the number of hidden layers for a multi-layer neural network [30].

The following independent variables are used for determining the regression model:

- Cluster number * Reference Distance
- Reference Distance
- The percentage of grouped items
- Quality measure for a neural network

The values for the first three independent variables are processed with clustering techniques and they are presented in table 1. The quality measure will take 1 or 0 depending on the Reference Distance and the architecture of neural network used by the NeuroShell program.

Based on the independent variables listed above a multiline regression model is constructed to determine the number of hidden layers for a multi-layer neural network having this form:

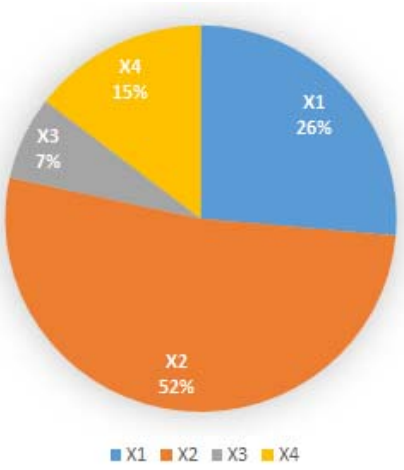
$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \quad (2)$$

In this model, the dependent variable y is the optimal number of hidden layers for multi-layer neural network and the significance of independent variables xi is given in table 2.

TABLE II. THE MEANING OF INDEPENDENT VARIABLES X_i USED TO DETERMINE THE NUMBER OF HIDDEN LAYERS

Independent variable	Meaning
X_1	Cluster number * Reference Distance
X_2	Reference Distance
X_3	The percentage of grouped items
X_4	Quality measure for a neural network.

To illustrate graphically the influence of the independent variables on the dependent variable, Fig. 3 shows the absolute values of partial correlation coefficients and the percentage contribution of each independent variable in the regression model.



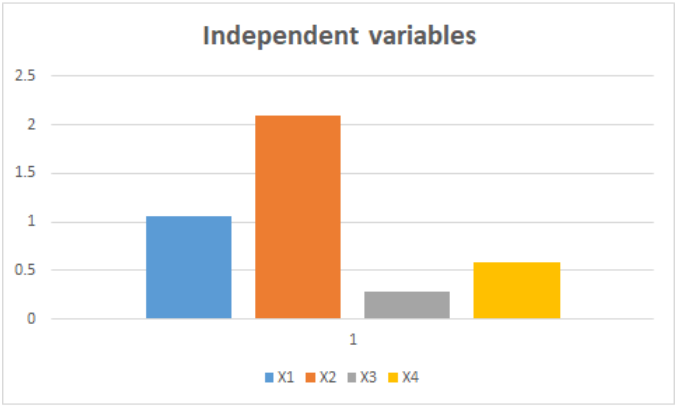


Fig. 3. Coefficients diagram for partial correlation in absolute values, indicating the contribution of each independent variable in the regression model to determine the number hidden layers.

The optimal number of hidden layers of a multi-layer neural network is influenced by several factors, the most important being the Reference Distance. The mathematical model constructed in the study of four factors, three of which are obtained using the clustering program and the other is the Quality measure characterizing the structure of the neural network.

B. Factors that determine the number of neurons in the hidden layers of a multi-layer neural network

Regression methods were used to determine the number of neurons in each hidden layer of neural network using the experimental data obtained previously by clustering program. Based on the experimental results obtained previously and in accordance with the above considerations, the following independent variables in determining the regression model for calculating the number of hidden neurons are considered:

- Number of inputs in the neural network
- Number of hidden layers, clusters
- The Reference Distance
- Quality measure for a neural network.

The first independent variable is the number of features uniquely characterizing input forms. The second variable is the number of clusters obtained by clustering programs used previously. The other two independent variables are as defined previously. With the factors listed above, a multiline regression model is built having this form:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \quad (3)$$

In the model, the dependent variable y is the number of hidden neurons of a multi-layer neural network. If there are more than one hidden layer, the number of hidden neurons will be divided equally by the number of hidden layers. Therefore, the hidden layers will always have the same number of neurons.

The significance of the independent variables xi is given in Table 3.

TABLE III. THE MEANING OF INDEPENDENT VARIABLES XI USED TO DETERMINE THE NUMBER OF HIDDEN NEURONS.

Independent variable	Meaning
X_1	Number of inputs in neural network
X_2	Number of hidden layers, groups
X_3	Reference distance
X_4	Quality measure for a neural network

To illustrate graphically the influence of the independent variables on the dependent variable, Fig. 4 shows the absolute values of partial correlation coefficients and the percentage contribution of each independent variable in the regression model.

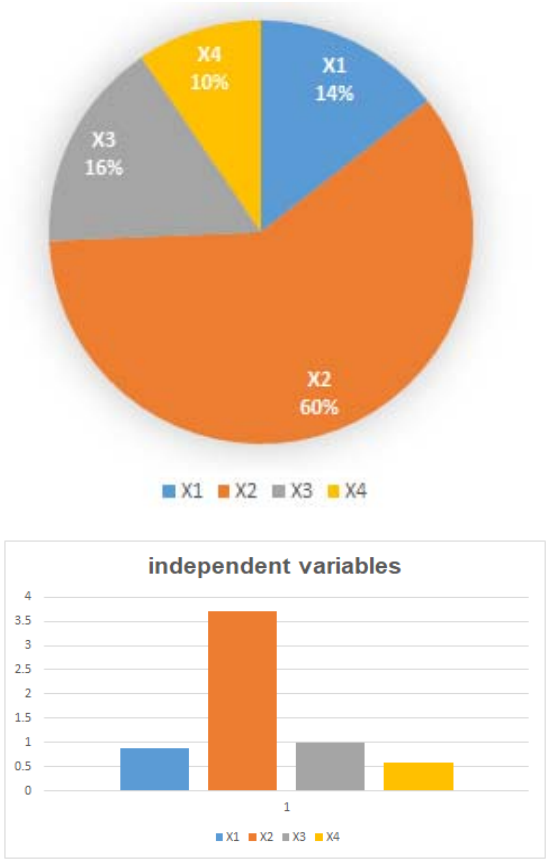


Fig. 4. Diagram of partial correlation coefficients in absolute values and the percentage of contribution of each independent variable in the regression model to determine the number of hidden neurons.

The importance of each parameter in determining the number of hidden neurons for a multi-layer neural network is easily seen.

The number of neurons in the hidden layers of a multi-layer neural network is influenced by several factors, the most important being the Reference Distance. The number of hidden layers and the number of inputs in a multi-layer neural network have influence in the number of neurons in hidden layers as well.

C. *Influence of independent variables on the number of hidden layers of a neural network*

Factors used to determine the number of hidden layers of a neural network have uneven influence. The importance of each factor compared to others is illustrated by:

The independent variables x_1 and x_2 have the most important influence compared to other independent variables with percentage of 78%. The Reference Distance variable x_2 has higher influence on the dependent variable by 52% followed by the independent variable x_1, which takes into account the number of groups and the Reference Distance, this variable has an important influence of 26%. The independent variable x_3 and x_4 have the lowest influence on the dependent variable y.

As mentioned earlier, the stability of the process of grouping must be taken into consideration in which two cases should be avoided. The first case presents a very short value of Reference Distance for which each element represents a group. The second case presents a relatively large value of Reference Distance to all the elements placed in one group. The Reference Distance chosen for the grouping affects directly the number of hidden layers of a neural network.

The independent variable x3 slightly influences on the dependent variable by 7% leading to the increase of the number of hidden layers.

The quality measure factor x4 is set to 1 for an architecture that has connections among hidden layers, and is set to 0 for an architecture without connections. This factor has a relatively low influence, from which can be concluded that the decrease of the number of connections between hidden layers causes the increase of the number of hidden layers. According to that, it appears that the degree of interconnections influences on the number of hidden layers.

The noise, incomplete records and those records showing large dissimilarities to other data cause an increase in the number of hidden layers.

The Multiple Determination Coefficient R^2 expresses the level of variation of the dependent variable y based on the variation of x. The percentage of variation of x is $R^2 = 92.48\%$ from the variation of the independent variable. The number of the training example, the number of dispersed elements from the training example, and the number of inputs and outputs of a neural network are influenced by 7.52%.

VI. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of data mining methods presented in this paper, the dataset will be tested with different neural network architectures. The best architecture in terms of error values will be compared with the architecture defined by data mining methods described above.

The simulation program used to determine the effectiveness of these methods is “Neuroshell ver. 2.0” of Ward Systems Group company program [32].

A. Experimental results for the number of hidden layers

For data analysis, experiments are used using different neural network architectures.

The following three types of neural network architectures are used:

- Neural network architecture with one hidden layer.
- Neural network architecture with two hidden layers.
- Neural network architecture with three hidden layers.

Fig. 5 shows the structure of three types of neural network architectures used.

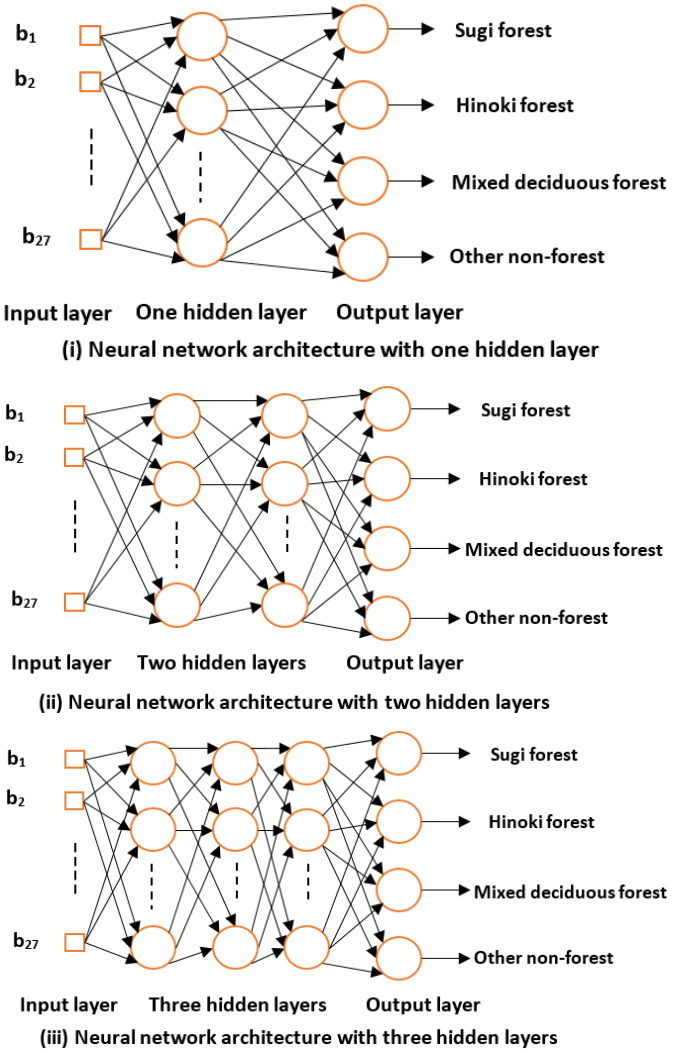


Fig. 5. Different neural network architectures with one, two and three hidden layers.

Fig. 6 presents error comparison for Forest Type Mapping dataset obtained by different neural network architectures.

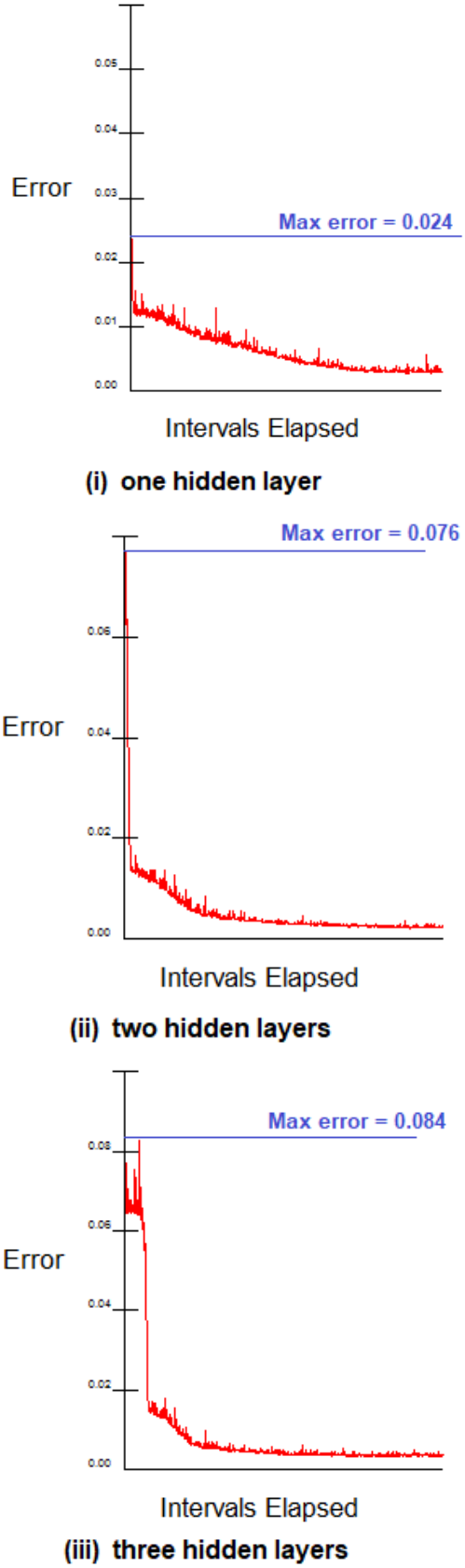


Fig. 6. The relative error results obtained by neural network architectures with one, two and three hidden layers for Forest Type Mapping dataset.

Please note that a network with two hidden layers has an error greater than 0.076 and for a network with three hidden layers has a largest error value of 0.084. A neural network with one hidden layer has a largest error value of 0.024.

From the above figure, it is seen quite clearly that the smallest error is obtained from a neural network with one hidden layer.

Therefore, the optimal neural network for classification of Forest Type Mapping dataset is the neural network architecture with one hidden layer. These results confirm the results obtained previously by grouping program.

The Indicator of Multiple Determination Coefficient (ratio) R^2 results using Neuroshell program:

- $R^2 = 0.9838$ for a neural network with one hidden layer.
- $R^2 = 0.9737$ for a neural network with two hidden layers.
- $R^2 = 0.9768$ for a neural network with three hidden layers.

The neural network with one hidden layer has the best result of multiple determination coefficient (ratio) R^2.

A comparison of the regression method results is conducted with relative error for the Forest Type Mapping dataset obtained previously by neural network architectures with different numbers of hidden layers.

In order to analyze a Forest Type Mapping dataset based on results of the grouping operation, the following values are obtained: $x_1 = 1*8.885$, $x_2 = 8.885$, $x_3 = 91.77$, $x_4 = 0$. Using these values in (2) give these results:

$$Y = (1.06) * x_1 + (2.08) * x_2 + (-0.28) * x_3 + (-0.58) * x_4 - 1.28$$

$Y = 0.904495 \approx 1$, this concludes that a neural network analyzing a Forest Type Mapping dataset should contain one hidden layer. This confirms the results of the best number of layers obtained by comparing error values of different neural network architectures.

Based on the analysis of experimental results obtained in this section, it can be confirmed that the optimal number of hidden layers of a multi-layer neural network can be determined using the regression method for the analyzed dataset as presented previously.

B. Experimental results for the number of hidden neurons

The Forest Type Mapping dataset will be taken as an example to demonstrate the differences between error results for each neural network.

Fig. 7 presents a comparison of Forest Type Mapping dataset errors obtained by neural network architectures with one hidden layer and different numbers of hidden neurons.

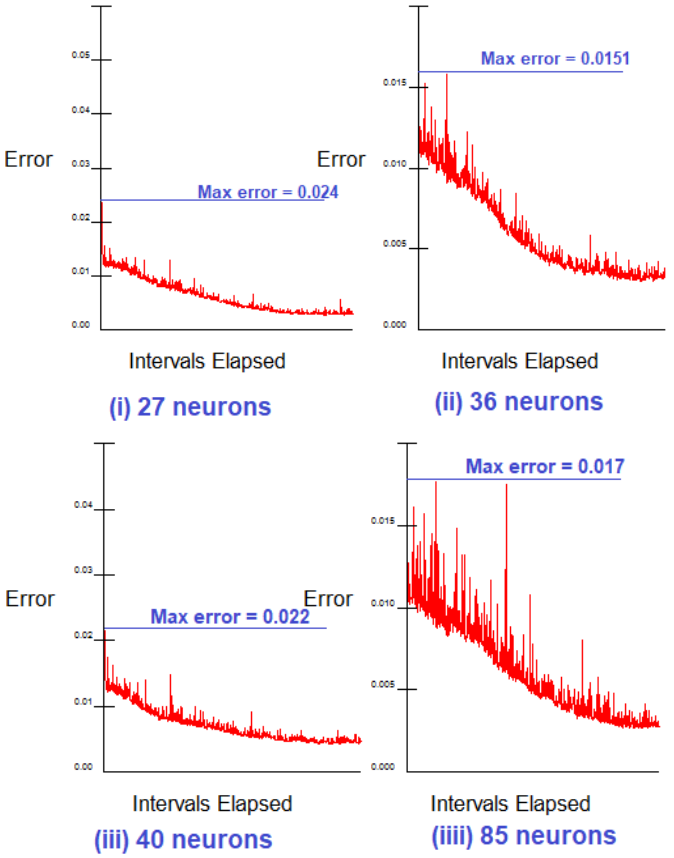


Fig. 7. The relative error for the Forest Type Mapping dataset obtained by neural network architectures with different numbers of hidden neurons.

Please note that the results of a neural network with 27 neurons has the largest number of errors with a value greater than 0.024 and for a neural network with 36 neurons has a larger error with a value less than 0.0151. For a neural network with 40 or 85 neurons, it has errors with a value greater than 0.022

The experimental results for Forest Type Mapping dataset make us conclude through error comparison obtained by different neural network architecture that a neural network with 36 hidden neurons has an error less than a neural network with less or greater number of hidden neurons.

A comparison of the regression method result with relative error for the Forest Type Mapping dataset obtained previously by neural network architectures with different numbers of hidden neurons will be conducted.

In order to analyze a Forest Type Mapping dataset, the following values are obtained as a result of the grouping operation: $x_1 = 27$, $x_2 = 1$, $x_3 = 8.885$, $x_4 = 0$. Using these values in (3) give these results:

$$Y = (0.886) * x_1 + (3.698) * x_2 + (0.994) * x_3 + (0.588) * x_4$$

$Y = 36.4704325 \approx 36$, this concludes that a neural network analyzing a Forest Type Mapping dataset should contain 36 hidden neurons. This confirms the results of the best number of hidden neurons of the neural network determined previously

using comparison of errors obtained by neural network architectures with different numbers of hidden neurons.

Based on the analysis of experimental results obtained in this section, it can be confirmed that the optimal number of hidden neurons in a multi-layer neural network can be determined using the regression method for the analyzed dataset as presented previously.

VII. CONCLUSIONS

Reference Distance has a high effect on determining the number of hidden layers of a multi-layer neural network compared to other factors when using clustering techniques to determine the number of hidden layers. The accuracy of determining the value of Reference Distance is essential.

It is seen through this paper that Pattern Recognition has a significant role in determining the optimal number of hidden layers of a multi-layer neural network.

The analysis of results carried out on the analyzed data in this study make us believe that the regression model built for calculating the number of hidden neurons required for a multi-layer neural network is viable. The model constructed can be used in practical applications and can at least provide information about the initial number of hidden neurons for a neural network. Based on the data from the training set, the user can then change this number up or down.

A new design strategy is proposed and developed for a neural network architecture, which makes the design of an optimal neural network unsupervised and decreases its building time.

REFERENCES

[1] J. Hirel, P. Gaussier and M. Quoy, "Biologically inspired neural networks for spatio-temporal planning in robotic navigation tasks," IEEE International Conference on Robotics and Biomimetics, pp. 1627-1632, 2011.

[2] H. Guanshan, "Neural network applications in sensor fusion for a mobile robot motion," WASE International Conference on Information Engineering, vol. 1, pp. 46-49, Aug 2010.

[3] A. Ansari, A. Abu Bakar, "A Comparative study of three artificial intelligence techniques: genetic algorithm, neural network, and fuzzy logic, on scheduling problem," 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, Kota Kinabalu, pp. 31-36, 2014.

[4] J.-Y. Wu, "An evolutionary multi-layer perceptron neural network for solving unconstrained global optimization problems," IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, pp. 1-6, 2016.

[5] K. Chen, S. Yang and C. Batur, "Effect of multi-hidden-layer structure on performance of BP neural network: Probe," Eighth International Conference on Natural Computation (ICNC), Chongqing, pp. 1-5, 2012.

[6] Y. Zhao, J. Chen and H. Vincent Poor, "Efficient neural network architecture for topology identification in smart grid," IEEE Global Conference on Signal and Information Processing (GlobalSIP), Washington, DC, pp. 811-815, 2016.

[7] R. Lovassy, L. T. Kóczy and L. Gál, "Fuzzy neural networks stability in terms of the number of hidden layers," IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, pp. 323-328, 2011.

[8] K. Shibata, Y. Ikeda, "Effect of number of hidden neurons on learning in large-scale layered neural networks," ICCAS-SICE, Fukuoka, pp. 5008-5013, 2009.

[9] S. Bunjongjit, A. Ngaopitakkul, C. Pothisarn and C. Jettanasen, "Improvement to reduce training time of back-propagation neural networks for discrimination between external short circuit and internal winding fault," International Conference on Information Science, Electronics and Electrical Engineering, Sapporo, pp. 614-618, 2014.

[10] Metin N.Gurcan, Heang-PingChan, BerkmanSahiner, LubomirHadjiiski, NicholasPetrick, Mark A.Helvie, "Optimal neural network architecture selection: improvement in computerized detection of microcalcifications," Academic Radiology Vol. 9, no. 4, pp. 420-429, April 2002.

[11] Devika Chhachhiya, Amita Sharma, Manish Gupta, "Designing optimal architecture of neural network with particle swarm optimization techniques specifically for educational dataset," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, pp. 52-57, 2017.

[12] C. A. Doukim, J. A. Dargham and A. Chekima , "Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique," 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), Kuala Lumpur, pp. 606-609, 2010.

[13] Chen-Yu Lin, Yung-Chih Chen, Chun-Yao Wang, Ching-Yi Huang and Chiou-Ting Hsu, "Minimization of number of neurons in voronoi diagram-based artificial neural networks," IEEE Transactions on Multi-Scale Computing Systems, vol. 2, no. 4, pp. 225-233, 2016.

[14] S. John Peter and S. P. Victor, "Cluster validity with minimum spanning tree based clustering," Journal of Theoretical and Applied Information Technology, vol. 17. no. 2, pp. 89-96, 2010.

[15] D. Greco, R. Tagliaferri and A. Serra, "Impact of different metrics on multi-view clustering," International Joint Conference on Neural Networks (IJCNN), Killarney, pp. 1-8, 2015.

[16] C. Guada, D. Gómez and J. Tinguaro Rodríguez, "Fuzzy image segmentation based on the hierarchical divide and link clustering algorithm," 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pp. 12-17, 2015.

[17] Grady, Leo, Schwartz and L. Eric, "Isoperimetric graph partitioning for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 3, pp. 469-475, March 2006.

[18] George Karypis, Eui-Hong Han, Vipin Kumar, "Chameleon: hierarchical clustering using dynamic modeling," Computer, vol. 32, no. 8, pp. 68-75, Aug 1999.

[19] Fionn Murtagh, Pedro Contreras, "Algorithms for hierarchical clustering: an overview, II," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 7, no. 6, 2017.

[20] Hussain Abu Dalbouh, Norita Md. Norwawi, "Improvement on agglomerative hierarchical clustering algorithm based on tree data structure with bidirectional approach," Intelligent Systems, Modelling and Simulation (ISMS), Third International Conference on Intelligent Systems Modelling and Simulation, pp. 25-30, 2012.

[21] Charu C. Aggarwal, Chandan K. Reddy, Data Clustering: Algorithms and Applications, by Chapman and Hall/CRC, 2013.

[22] I. Gath, Amir B. Geva, "Unsupervised optimal fuzzy clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 773-780, Jul 1989.

[23] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, 2006.

[24] Le Xu; Mo-Yuen Chow, "Power distribution systems fault cause identification using logistic regression and artificial neural network," Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems, Arlington, VA, 2005.

[25] Jeng-Neng Hwang , Shyh-Rong Lay, M. Maechler, "Regression modeling in back-propagation and projection pursuit learning," IEEE Transactions on Neural Networks, vol. 5, no. 3, pp. 342-353, May 1994.

[26] Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[27] 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, vol. 14, no. 6, pp. 739-745, Dec 2009.

[28] “Forest Type Mapping Data Set” UCI Machine Learning Repository. Irvine, California: University of California, School of Information and Computer Science, [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Forest+type+mapping

[29] "Irvine machine learning dataset and benchmark," Center for Machine Learning and Intelligent Systems, [Online]. Available: http://archive.ics.uci.edu/ml/datasets.html

[30] 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," IEEE 5th International Conference on Cybernetics and Intelligent Systems (CIS), Qingdao, pp. 374-379, 2011.

[31] Elliott, A. C. & Woodward, W. A, Statistical Analysis Quick Reference Guidebook, SAGE Publications Ltd, Thousand Oaks, CA, 2007

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