Proceedings of the *International Conference on Theory and Application of Mathematics and Informatics ICTAMI 2005* - Alba Iulia, Romania

AN EXPERIMENTAL SPEAKER-INDEPENDENT SYSTEM FOR ISOLATED WORD RECOGNITION IMPLEMENTED FOR ROMANIAN LANGUAGE

Marieta Gâta

Abstract. The research presented in this paper is based on the artificial neural networks recognition paradigm applied to Romanian isolated word recognition. The network, which is composed by three layer (a Multilayer Perceptron), is trained by conventional Back-propagation algorithm. The ANN speech recognition system based on Mel Frequency Cepstral Coefficients was developed using Matlab toolkit. The system was tested on the 60 isolated word Romanian speech corpus and it was training on a 60 distinct isolated word recognition task. The word recognition accuracy obtained was about 81.6 %.

Here we presented a set of experiments involving artificial neural networks used in a speech recognizer system. The goal was to test our program for different speakers and to optimize a number of factors like: structures of the network, input window size, network training procedure, etc. All this factors can influence the performance of our system. The platform for research and development of our recognition system is using a program that was developed in Matlab.

2000 *Mathematics Subject Classification*: 68T10, 82C32, 62M45.

1. Implementation of an artificial neural network for speech recognition

The paper presents an application that implements a type of the artificial neural network: MLP (Multi Layer Perceptron). This network is based on Backpropagation learn, which is the adjustment of the weights according to the global error of the network. This adjustment is done in backward order of processing.

2. Structure of our speech recognition system

The aim of this research was to develop a baseline small vocabulary isolated word Romanians ANN based speech recognition system.

The ANN system has already been successfully applied in connected and continuous English digits recognition tasks. In case of recognition systems developed for Romanian language we are still at the beginning, but some results were obtained for isolated word recognition [2] and it have been made some tentative for continuous speech recognition [3].

In the next sections we present some results obtained by our team for an isolated words recognition system developed by ourselves.

The research presented here is new because it is based on the study upon learning rate, number of epochs, training time, number of neurons in hidden layer applied to Romanian isolated word recognition for our condition: database with 120 records, closed room and neural network described bellow.

The platform for research and development of our recognition system is using a program made by us, which was developed in Matlab R12.

In our experiments we used a feed-forward neural network, the network was trained with Backpropagation algorithm. This MLP network has 150 neurons on the input layer, 100 neurons on the hidden layer and 12 on the output layer. If in the network we add another intermediate layer this will affect the calculus time, which can increase more that without this layer.

Our database is formed by ten speakers, each speaker record twelve different words, in this case the months of the year, the type of recording for each file will be wav format. A half of the utterance (wav files) was used in the training step and the other half in the testing step of the recognition process. As a training function we use tansig in the hidden layer and logsig in the output layer.

The values of the threshold assure a natural generalization of the net. The constraint of very well learning of a set of examples it goes to the errors of recognition through particularization and locking of the certain and rigid set of training. Therefore the training must be done until is achieved a certain value of the error, named critical value, below which the network doesn't generalize, and over which the networks doesn't offer proper answers.

3. Experimental results

In this paper we achieved an application for speech recognition. We conducted a series of experiments using this application for different words and

different speakers. As a base for our result we use twelve Romanian words: the name of the months, from January to December, in Romanian (Ianuarie, Februarie, Martie, Aprilie, Mai, Iunie, Iulie, August, Septembrie, Octombrie, Noiembrie, Decembrie).

Figure 1 shows an example of "Ianuarie" word recorded by the program as wav file. Figure 2 presents spectral component (the spectrogram) of the same word "Ianuarie" uttered by the same speaker as in Figure 1.

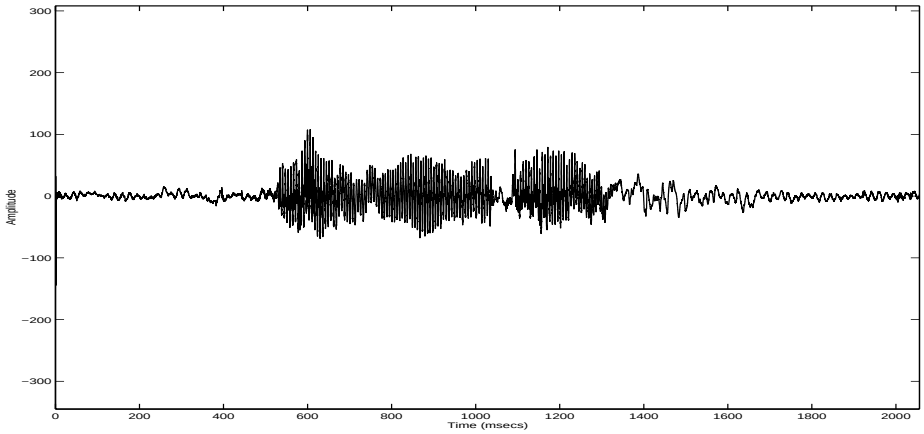

Figure 1

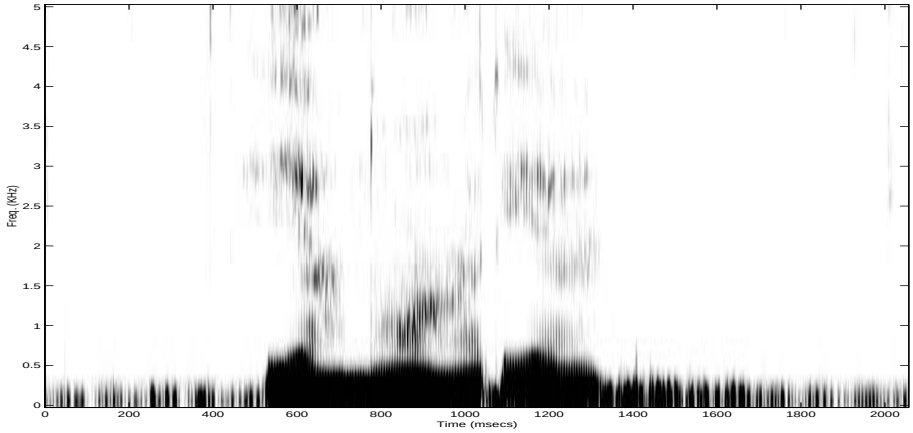

Figure 2

Our contribution is the study upon learning rate, number of epochs, training time, number of neurons in hidden layer in speech recognition for our condition: database with 120 records, closed room and neural network described before.

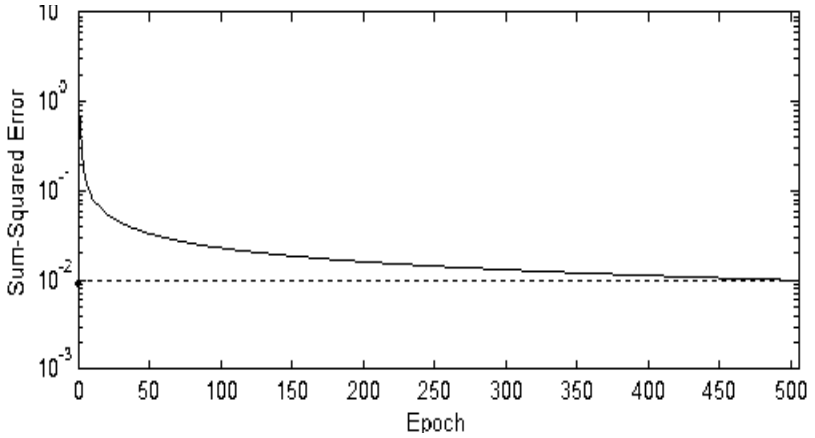

Figure 3

Figure 4 shows recognition rate for twelve Romanian word, number of epochs 2000, learning rate 0.1, and sum-squared error 0.01. Figure 3 presents an example of Sum-Squared Network Error for 2000 epochs for the same experimental dates as in Figure 4.

Figure 4

We obtained several results in this paper, which are presented in Table 1 and Table 2:

	Number of epochs (for learning rate=0.1)								
	1000	2000	3000	4000	5000	6000	7000	8000	9000
Recognition rate(%)	78%	81%	78%	76%	74%	71%	64%	61%	53%
Training time(min)	1	2	2.5	3	3.5	4	6	7	9

Table 1

	Learning rate (for epochs=5000)								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Recognition rate(%)	81%	74%	57%	48%	38%	32%	26%	24%	21%
Training time(min)	4	4.5	5	6.5	7	7.5	8	8.5	9

Table 2

Another parameters that we try to change was the characteristics of the neural network, which implies:

- number and type of inputs (which means the choice of the features for the pattern-classification system)
- connectivity of the network (this characteristic involves the size of the network -the number of hidden layers and the number of nodes in each layer between input and output- There is no rule as to how large or small such hidden layers must be)
- offset/threshold (this choice is a part of the training procedure, which chooses values for the interconnection weights w_{ij} and the offset Φ)

About these properties of the network we find a few observations.

Connectivity of the network: if the hidden layers are large, then it will be difficult to train the network (there will be too many parameters to estimate) and if the hidden layers are too small, the network may not be able to precisely classify all the desired input patterns. Usually practical systems must balance through these two competing effects.

Nonlinearity: the exact choice of the nonlinearity, f, is not very important in terms of the network performance. It must be continuous and differentiable for the training algorithm to be applicable.

4. Conclusions and discussions

Our observation is that the rate of recognition is decreasing with number of epochs. Other observation is that if learning rate is increasing than recognition rate is decreasing. Regarding the training time we observe that this is growing with the number of epochs and also is growing with learning rate usually.

The best recognition rate obtained by us is 81 %. We obtained the same result even if we decreased learning rate to 0.01 and number of epochs remains the same, 2000.

Regarding the number of neurons in the hidden layer, in our experiments we obtained as the best value 100 (depending of obtained results and calculus time). In our tests this value was between the values 60 and 150.

Also we observed that if learning rate is high then the algorithm is unstable and if learning rate is small then the number of steps for convergence is high.

5. Future work

In this experiments we are at the beginning of our experiments that developed a speech recognition system based on the neural networks. In future experiments we will try to improve few parameters which can influence the performance of the system. Also, the future experiments should be extended to a much bigger speech databases and a large number of speaker.

Here we present a number of experiments involving neural networks used in speech recognition. Our target was to test our software -that we have implemented- to study the influence of a few parameters -such as learning rate, number of epochs, training time, number of neurons in hidden layer- in speech recognition for our condition of work. In our future tests we try to record and process for a speech recognition system for Romanian language not only for isolated word but also for connected word.

References

[1] Cole, R. A. Cole, Mariani, J., Uszkoreit, H., Zaenen, A., Zue, V., *Survey of the State of the Art in Human Language Technology*, Cambridge University Press, 1996

[2] Giurgiu, M., *Contributions to the Isolated Word Recognition for Romanian Language*, (in Romanian) PhD Thesis, TUCN, 1996

[3] Giurgiu, M., *Experimental Informations Retrieval System Based on Romanian Continuous Speech Recognition*, 2nd Conference on Speech Technology and Human Computer Dialogue, SpeD 2003, Bucharest, 161-166, April 10-11, 2003

[4] Green, P., Renals, S., *Speech Technology*, 2002

[5] Huang, X., Acero, A., Hon, H. W., Reddy, R., *Spoken Language Processing: A Guide to Theory, Algorithm & System Development*, 2001, Prentice Hall PTR

[6] Hagan, M. T., H. B. Demuth, and M. H. Beale, *Neural Network Design*, 1996, Boston, MA PWS Publishing

[7] Lippmann, R., *An Introduction to Computing with Neural Nets*, IEEE ASSP Mag., 4(2):4-22, April 1987

[8] Rabiner, R L., Juang, B.H., *Fundamentals of Speech Recognition*, Prentice Hall, New Jersey, USA, 1993

[9] Toderean, G., Costeiu, M., Giurgiu, M., *Artificial Neural Networks* (in Romanian), Microinformatica Cluj-Napoca, 1995

[10] Weibel, A., Hanazawa, T., Hinton, G., Shikano K., Lang, K.J., *Phoneme Recognition Using Time Delay Neural Networks*, IEEE Trans. Acoustics Speech Signal Processing, ASSP-37: 328-339, 1989

Marieta Gâta
Department of Mathematics and Computer Science
North University of Baia Mare
Address Victoriei, No. 76, 430144, Baia Mare, Romania
e-mail:*marietag@ubm.ro*