

**A LEARNING PROCESS OF MULTILAYER
PERCEPTRON FOR SPEECH RECOGNITION**

Norelhouda Azzizi¹, Abdelouhab Zaatri² §

¹Department of Mathematics
Faculty of Exact Sciences
University of Freres Mentouri
Constantine, ALGERIA

²Department of Mechanical Engineering
Faculty of Engineering Sciences
University of Freres Mentouri
Constantine, ALGERIA

Abstract: For learning artificial systems as well as for living systems, it is generally proven that the learning performances improve with the experience. This paper seeks to analyze the learning process of an artificial system: a Multi-Layer Perceptron Neural Nets (MLP-NN) used for word recognition and dedicated for robot control. As the MLP requires references for the spoken words, we have provided these references by means of a supervised classifier based on minimizing the mean square error.

We are particularly interested by estimating the minimal number of trials required to ensure the recognition of some spoken words by the MLP-NN with an acceptable predefined error. To this purpose, we have experimentally performed the learning process of the recognition of some specific words. For each word, we have recorded the performance improvement with respect to the number of trials enabling to draw the learning curve. The mathematical modeling of these curves presents a bi-exponential law profile while the mathematical modeling of human performance show generally a power law profile. The obtained results have led to a better understanding of the artificial system performance under the influence of internal

Received: February 22, 2016

Published: May 7, 2016

© 2016 Academic Publications, Ltd.

url: www.acadpubl.eu

§Correspondence author

and external human and technological factors.

AMS Subject Classification: 00A05

Key Words: learning curve, supervised learning, MLP, mathematical modeling, neural networks, word recognition

1. Introduction

The learning process is involved in various research domains such as neuroscience, psychology, education, training, control of artificial systems, industry, etc. Understanding the learning process for living systems as well as artificial systems is a crucial issue. For living systems, this may help to reduce the learning duration both for normal and disabled people. For artificial systems, this may help to minimize the learning duration and to control the effects of external and internal factors that may influence the process.

The learning process is a dynamical one. However, dynamic estimation methods are not so often used because it may depend on many factors and the consensus is hard to establish about when the learning has effectively occurred [1]. For this purpose, experimental studies are adopted for analyzing the performance of the learning process. They lead to build up the learning curves which are obtained by drawing the performance of the learning process with respect to the number of trials. In many cases, reinforcement can be introduced in order to empower the learning process [1]. The form of the learning curve can provide information about the learning process. The duration of the dynamical zone, the abruptness of the slope, and the asymptotic rate. These elements can provide a qualitative analysis of the phenomenon with respect to different parameters [1,2].

For our application; we are concerned by the analysis of a supervised learning artificial system dedicated for the recognition of isolated words [3]. However, because of the complex nature of voice signal, the speech recognition still remains a hard issue. Most speech recognition systems use a learning process to identify the correct response of a spoken command. In this context, neural nets models can be used for estimating the output of nonlinear systems in the case of noisy and sensitive process to various parameters. There exist different methods for speech recognition of isolated words [4,5,6,7,8].

In our research, we aim to analyze the learning profile of a supervised MLP-NN used for word recognition and dedicated for robot control [3]. We intend to model the profile of the learning curve for the given spoken words. We,

experimentally, test the evolution of the performance of the learning process by determining the minimal number of trials to obtain the recognition of a spoken word with an acceptable error. Some elementary results of human operator concerning the learning process have been also initiated and compared with our results.

2. MLP Supervised Learning for Word Recognition

The principle used for most word recognition systems comprises two phases: the learning phase and the recognition phase. The learning phase consists of creating a list of words which are stored into a dictionary as reference words. The recognition phase consists of identifying if possible a spoken unknown word to one of the reference words.

We have used a word recognition system which has been described in [3]. The estimation of the reference words (robot commands) are obtained by a supervised classifier based on the minimization of the mean square error. These reference words are stored into the dictionary and used by the MLP for comparison with a new pronounced word [3]. The role of the MLP classifier is to select the most similar reference word with respect to an unknown word. The choice is based on the calculation of the distance between the unknown word characterized by its cepstral coefficients a_{ij}^* and all the reference words a_{ij} (nearest neighbor). Practically, the system has been tested on a dictionary of four commands (START, STOP, UP, DOWN).

The implementation of the MLP was carried out by using the NN toolbox of Matlab software [3]. It is composed of an input layer and an output layer including one hidden layer in between (Figure 1). The input data of the MLP are the MFCC which are recorded into a file in a form of a matrix named "sepstr.mat". The MLP uses 12×32 neurons for the input layer. The reference word was determined from the previous learning process. A supervised training was adopted comparing actual spoken words with those stored on the dictionary. After the achievement of the learning process, Matlab provided automatically the hidden layer constituted of 32 neurons. The output layer is constituted of 4 neurons which corresponds to the reference words stored on the dictionary (START, STOP, UP, DOWN).

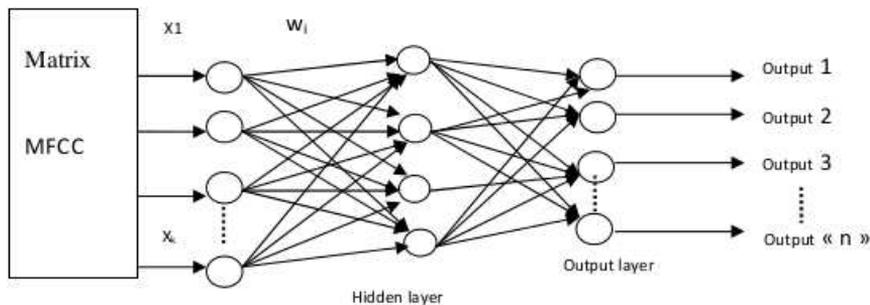


Figure 1: MLP classifier

3. Experimental Results

The learning phase for the MLP was derived as follows. For each spoken word, we use a certain number of trials (N). We store the obtained performance with respect to the number of trials until the learning process converges towards some limited error. The adopted performance measure $mse(N)$ is the mean square error of MFCC which is computed with respect to the number of trials N as:

$$mse(N) = \left(\sum (a_{ij} - a_{ij})^2 \right) / N.$$

For each word command, we have performed experiments to obtain the corresponding learning curve [3]. We present graphically an example illustrating the learning curve concerning the word command UP (Figure 2). The experimental learning curve is constituted by the blue dots while the red curve is one of its approximation curve.

The learning curve can be defined for human operator as well as for automated learning systems. Some of our elementary results concerning the learning process of human operator involving his/her sensory systems have been also initiated in a previous work [9] and presented in Figure 3.

It shows the decreasing rate of the learning process: time gained w.r.t the number of trials.

4. Mathematical Modeling of Learning Curves

In general, the performance of a learning system improves with the number of trials. However, various models of learning curves have been proposed: expo-

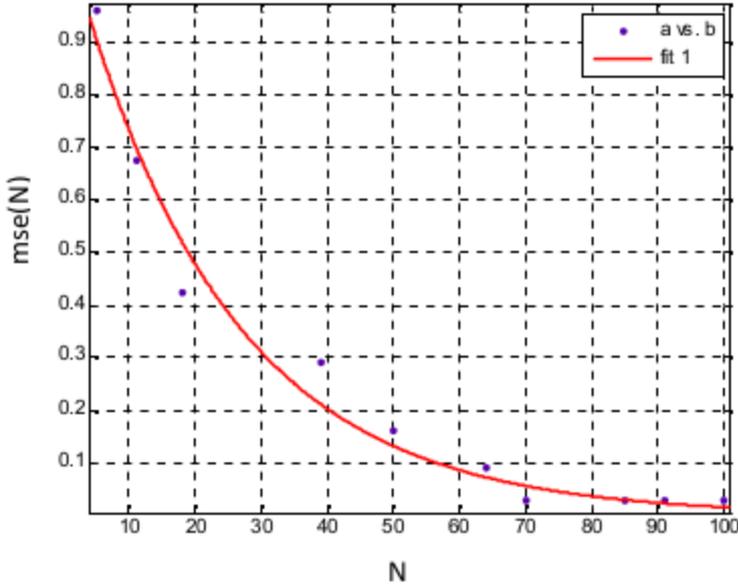


Figure 2: mse(N) with the number of trials for the word UP

nential growth, exponential rise or fall to a limit, power law [10,14,15]. They are common and practical softwares that can be used to analyze learning curves such as Excel, Excel Om and POM.

In our experiments, we have tested various approximation functions for our spoken words in order to obtain an appropriate model of the learning curve. As a result, we have noticed that the most appropriate approximation is a bi-exponential function. The experiments concerning the word command UP show that the mse(N) decreases with the number of trials; improving therefore the performance of the learning process (Figure 2). The bi-exponential approximation function is given by the Curve Fitting toolbox of Matlab software as:

$$mse(N) = a * \exp(b * N) + c * \exp(d * N),$$

with the following coefficients:

$$a = 2.019e + 004, \quad b = -0.0634, \quad c = -2.019e + 004, \quad d = -0.0634.$$

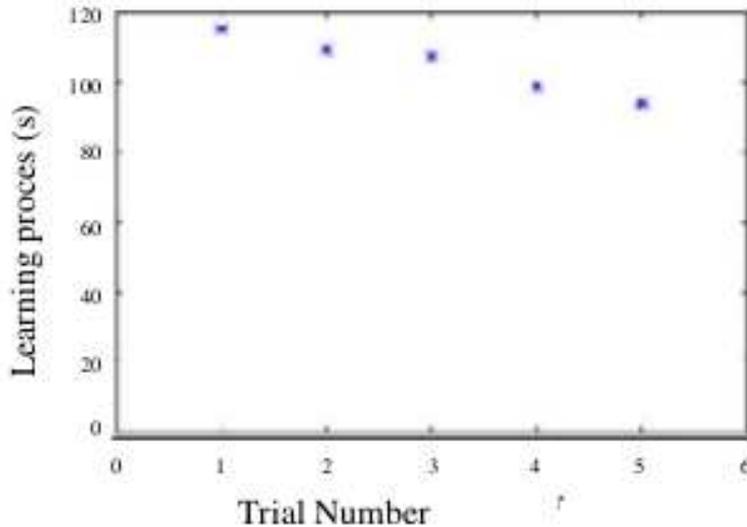


Figure 3: Learning process for human recognition

More learning curves concerning other word commands with their approximation functions can be found in [3].

Similarly, some authors have also applied NN for words recognition and analyzed the learning process with respect to some factors such as the number of hidden layers. They have reported some learning curves: the mean square error versus the trial number. Their results have confirmed that the global profile of the learning curve shows a decreasing error (increasing performance) with the trial number[11,12].

Some experimental analysis of human learning curve tested upon the hearing sensory systems have also been performed by [13,14]. An example of word recognition applied to children is reported in[13]. It concerns the short-term word learning rate in children with normal hearing and children with hearing loss. Other recent research concerning hearing loss with age based on an experimental study leading to learning curve are presented in [14].

As a general remarks, the obtained learning curves of living and artificial systems for word recognition are typical to each situation since it depends on the specificity of the system and on its environment. After testing experimentally

the MLP with some spoken robot commands (START, STOP, UP, DOWN), we have effectively determined the minimum number of trials required by the learning process to ensure some limited error. We have noticed that this number depends on the structure of the spoken word itself, on the speaker characteristics, on the used equipment and on the environment noise.

5. Conclusion

We presented an experimental technique to estimate the minimal number of trials for the learning phase to ensure an acceptable performance of a supervised neural network dedicated to word recognition used for robot commands.

We have found that the learning process tends to improve the performances of the word recognition in a way that the error decreases as a bi-exponential function with respect to the number of trials.

Finally, we have noticed that the success rate of the MLP and the minimal number of trials used for the learning process depend on the structure of the spoken word, on the used equipments, on the environment noise and on the speaker characteristics.

Some elementary qualitative results of human operator concerning the learning process have been also initiated and compared to the word recognition neural nets.

References

- [1] C. R. Pittman, S. Fairhurst, P. Balsam, , *The learning curve: Implications of a quantitative analysis* , Proceedings of the national academy of sciences of the united states of america PNAS, vol. 101 (36), doi: 10.1073/pnas.0404965101 (2010), 785797.
- [2] R. L. S. Monteiro,, T.K. G. Carneiro, J. R. A. Fontoura, V. L. da Silva, M. A. Moret , *A Model for Improving the Learning Curves of Artificial Neural Networks* , Hernane Borges de Barros Pereira, <http://dx.doi.org/10.1371/journal.pone.0149874> (2016).
- [3] A.Zaatri, N.Azzizi, , *Design experiments for voice commands using neural networks*, vol 12 (3), world journal of engineering, doi:10.1260/1708-5284.11.4.441 (2015).
- [4] K. J. Lang,A. H. Waibel, *A Time-Delay Neural network Architecture for Isolated Word Recognition*, Neural Networks, vol. 3, doi: 10.1016/08936080(90)90044-L (1990).
- [5] L.R. Rabiner, *A tutorial on Hidden Markov Models and selected applications in Speech Recognition*,Proceedings of the IEEE Journal, vol. 77, Issue: 2, doi: 10.1109/5.18626. (1989).
- [6] D. Paul, R. Parekh, *Automated speech recognition of isolated words using neural networks*, International Journal of Engineering Science and Technology (IJEST), 3(6) (2011), 4993-5000.

- [7] K.I. Funahashi, Y. Nakamura, *Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks*, Neural Networks, vol. 6, doi: 10.1016/S0893-6080(05)80125-X (1993).
- [8] Pinto. J. P, *Multilayer Perceptron Based Hierarchical Acoustic Modeling for Automatic Speech Recognition*, These N.4649,Lausanne, EPFL (2010).
- [9] A. Zaatri, *Investigation into integrated supervised control systems*, Ph.D. Thesis, Katholieke Univesiteit Leuven, Belgium (2000).
- [10] F. E.Ritter, *The learning curve*, In International Encyclopedia of the Social and Behavioral Sciences, doi: 10.1016/B0-08-043076-7/01480-7 (2002), 8602-8605.
- [11]] W.Gevaert, G.Tsenov, V. Mladenov, *Neural Networks used for Speech Recognition* , Journal of Automatic Control, University of Belgrade, VOL. 20, doi:10.2298/JAC1001001G (2010), 1-7.
- [12] N.Srivastava, *Speech Recognition using Artificial Neural Network* , International Journal of Engineering Science and Innovative Technology (IJESIT), Volume 3, Issue 3 (2014).
- [13] B.J. Baars, N. M. Gage, *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, Academic Press, doi:10.1038/npre.2012.6775.1 (2010).
- [14] H.Thomas, C. Mathews, J. Ward, *Learning Curve Models of Construction Productivity*, J. Constr. Eng. Manage., 10.1061/(ASCE)0733-9364(1986)112:2(245), 245-258 (1986).
- [15] S. Smale, D. X. Zhou, *Estimating the approximation error in learning theory*, Analysis and Applications, vol. 01, Issue 01, doi: 10.1142/S0219530503000089 (2003).