NEURAL NETWORKS APPLIED A INTELLIGENT SYSTEM OF SPEECH RECOGNITION BASED ON DCT PARAMETRIC MODELS OF LOW ORDER

Priscila Rocha*, Washington Silva†, Allan Barros*

*Federal University of Maranhão
São Luís, Maranhão, Brazil

†Federal Institute of Maranhão
São Luís, Maranhão, Brazil

Emails: priscilalima_rocha@.hotmail.com, washington.silva@ifma.edu.br, akduailibe@gmail.com

Abstract— This paper proposes the development of a Numerical Command Recognition System of Speech Signal based on Neural Networks and DCT models. Thus, two configurations of neural networks, the Multilayer Perceptron (MLP) and Learning Vector Quantization (LVQ) are evaluated by their performance in speech signal recognition, whose encoding is made by the mel-cepstral coefficients that are used to generate a two-dimensional time matrix by Discrete Cosine Transform (DCT). The selection of the best configuration of neural network for classification of the patterns was carried out by comparative analysis of performance of the MLP and LVQ networks through training, validation and test of the network topology and learning algorithms previously established. For demonstration of the performance of the proposed analysis methodology, the obtained results were compared with other methods of classification given by Gaussian Mixture Models (GMM) and Support Vector Machines (SVM).

Keywords— Automatic Speech Recognition, Neural Network, Discrete Cosine Transform.

Resumo— Este artigo propõe o desenvolvimento de um Sistema de Reconhecimento de Comandos Numéricos de Sinal de Voz baseado nas configurações de redes neurais Perceptron de Múltiplas Camadas (MLP) e a Aprendizado por Quantização Vetorial (LVQ). A codificação do sinal de voz é feita por meio dos coeficientes mel-cepstrais que são usados para gerar uma matriz temporal bidimensional pela aplicação da Transformada Cosseno Discreta (TCD). Para demonstrar o desempenho da metodologia de análise proposta, os resultados obtidos foram comparados com outros métodos de classificação, dados pelos Modelos de Misturas Gaussianas (GMM) e Máquinas de Vetores de Suporte (SVM).

Palavras-chave— Reconhecimento Automático de Voz, Rede Neural, Transformada Cosseno Discreta.

1 Introduction

The extensive development of research in the speech processing area shows the effort to improve the performance of speech recognition systems for practical applications. The use of such systems allow autonomy in areas such as telephony, wherein service requests are directed through speech commands (Cardoso et al., 2010); in the automotive industry through the activation of devices inside vehicles; in computing systems by utility programs on computers, besides the application in robotics, residential and hospital automation for accessibility of people with locomotor and visual disease (A.Cubukcu et al., 2015).

Therefore, it is proposed in this paper the development of a Numerical Command Recognition System of Speech Signal with two configurations of neural networks, the Perceptron Multilayer and Learning Vector Quantization. They will be evaluated by its performances in speech signal recognition, whose encoding is done using the mel-cepstral coefficients (MFCC) and discrete cosine transform (DCT) (Silva e Serra, 2014). The mel-cepstral coefficients and the DCT are used for generating of a two-dimensional time matrix which represents, through a reduced set of parameters, the pattern of speech signal to be used as training, validation and testing input of the networks neural.

1.1 Proposal Analysis Methodology

The proposed speech recognition system in this paper aims to classify patterns of speech signals, represented by locutions from Brazilian Portuguese of the digit '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', through the recognizer characterized by the neural network. The adopted methodology uses a reduced number of parameters to represent each pattern obtained in the speech signal pre-processing stage for generating two-dimensional time matrices 2, 3 e 4. The time two-dimensional matrix reproduces the global and local variations in time, as well as the spectral envelope of the speech signal. The use of neural network as classifier in speech recognition system using the low-order parameters generated by the two-dimensional matrix of low temporal order and the comparison of the results are the main contributions of this work.

The analysis of performance between two configuration MLP and LVQ Neural Networks is carried out in two phases: first, the specified topologies the observation of the behavior of the MLP and LVQ networks in training and validation process and the selection of the topologies that accomplish global validation with hit above 80%. Then the selected topologies are tested with different parameters of the speech signal that are not used in the training process and the results of classifications of the MLP and LVQ neural networks are shown. The study performance of LVQ Neural Network, presented in this paper, applied in the speech recognition with low-order models provides an alternative approach to the MLP classifier that is the neural network most used in speech recognition problems (Haykin, 2001).

2 Speech Recognition System based on Networks Neural

It is shown in Figure 1 the block diagram of the proposed speech recognition system.

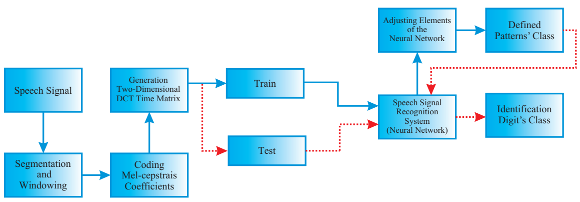

Figure 1: Block diagram of the Speech Recognition System using Neural Networks

2.1 Generation of two-dimensional DCT time matrix

After obtaining the mel-cepstral coefficients from the speech signal, encoding through discrete cosine transform (DCT) was carried out, which enables to synthesize the long-term variations of the spectral envelope of the speech signal. The result of this encoding was the generation of a two-dimensional DCT time matrix (Silva e Serra, 2014), obtained by (1):

$$C_k(n,T) = \frac{1}{N}\sum_{t=1}^{T} \mathrm{mfcc}_k(t)\cos\left[\frac{(2t-1)n\pi}{2T}\right] \quad (1)$$

where k, that varies of $1 \leq k \leq K$, is the k-th line component of t-th segment of the matrix. K is the number of mel-cepstral coefficient; n, that varies of $1 \leq n \leq N$, is the n-th column. n is the order of the matrix DCT; T is the number of vectors of observation of the mel-cepstral coefficients in time axis; $\mathrm{mfcc}_k(t)$ represent the mel-cepstral coefficients.

Thus, for each locution of the digit $\mathbf{D}$ to be recognized there is a two-dimensional DCT time matrix C_{kn}^{jm}, where $j = 0, 1, 2, \cdots, 9$ represent the digit to be encoded and $m = 0, 1, 2, \cdots, 9$ represent the example taken for each digit. These matrices were transformed in column vectors C_N^{jm}, where N is the number of parameters of the two-dimensional matrix C_{kn}^{jm}. The column vector C_N^{jm} preserves the temporal alignment of mel-cepstral coefficients and its general term is given by (2):

$$C_N^{jm} = [c_{11}^{jm}, c_{12}^{jm}, \cdots, c_{1n}^{jm} c_{21}^{jm}, c_{22}^{jm}, \cdots, c_{2n}^{jm}, \cdots, c_{kn}^{jm}]' \quad (2)$$

The vectors C_N^{jm} were used as input patterns of the neural network. The dimension of C_N^{jm} defines the number of input of the neural network. In order to compare the performance of the neural network when the number of parameters that compose the input patterns of speech is increased, two-dimensional matrices C_{kn}^{jm} were generated of order $n = 2, 3$ e 4, thereby obtaining the patterns to be recognized by neural networks represented by C_N^{jm}, where $N = 4$, 9 e 16, respectively.

2.2 Design of Neural Networks

The design of the neural networks is carried out through simulations of combinations of elements of the network topology and learning algorithms previously established and also based on obtained results in other similar works of patters classification. The configurations of MLP and LVQ network were used to the pattern recognition according to speech signal encoding by two-dimensional time DCT matrix. Then, in Table 1 and Table 2 present, respectively, the elements related to topology and training algorithms previously established for MLP and LVQ networks that were combined during the training phase.

Table 1: Training elements of MLP Neural Networks

Elements	Symbol	Defined Range
Training Algorithms	-	GD,GDM,RP,LM[1]
N° of Hidden Layers	-	1 e 2
N° of Hidden Neuron 1 Hidden Layer	n_1	20,25,30,35,40,45,50,55,60
N° of Hidden Neuron 2 Hidden Layers	n_1 e n_2	$n_2 = 15$
Learning Rate	η	0.01
Momentum Term	α	0.8
N° of Epoch	-	1000
Activation Function	-	Hyperbolic Tangent

Table 2: Training elements of LVQ Neural Networks

Elements	Symbol	Defined Range
Training Algorithms	-	LVQ-1
N° of Hidden Layers	n_1	20,25,30,35,40,45,50,55,60
Learning Rate	η	0.01
N° of Epoch	-	1000

It can be seen in Tables 1 and 2 that the LVQ network has a smaller number of elements that can be associated to the simulations of training. This is due to simplicity of the architecture and topology of the LVQ network. For MLP configuration networks with two hidden layers were simulated

[1]GD: Gradient Descent; GDM: Gradient Descent with Momentum; RP: Resilient Backpropagation; LM: Levenberg-Marquardt

in order to verify the need to increase the number of hidden layers to extract the features contained in the input patterns presented to the network.

During the research conducted for the preparation of this paper, it was observed that the initialization of the set of weights for the MLP neural network impacted the final results. Therefore, to realize this fact, four training (T_1, T_2, T_3, T_4) with different initializations of the set of weights made over a random uniform distribution between interval $[-0.01, 0.01]$ was carried out for each proposed topology for MLP network. Thus, it was possible to observe the behavior of neural networks in relation to training time and the ability to generalize, because appropriate set of initial weights allows a reduction in training time and a high probability of achieving the overall minimum of error function. In addition, this set can significantly improve performance in generalization.

2.3 Training and Testing Sets

The Training and testing of the MLP and LVQ Neural Networks were carried out with patterns of speech signals from selection of locutions from voice bank EPUSP, INATEL and IFMA. These sets are changed depending on the order of the two-dimensional matrix used to generate the patterns. The training and test sets are composed as follows:

1. Training Sets Ω_{NL}^{Tr}: This set represents the number of parameters of patterns of the set, where N = 4, 9, 16, L the total number of locutions and Tr indicates that is training set, is composed of 200 locutions, where half are female speakers and the other half are male speakers. The training set was partitioned in the estimation subset Ω_N^E (it contains 80% of the total of Ω_{NL}^{Tr}) and in the validation subset Ω_N^V (it contains 20% of the total of training set) ($\Omega_{N200}^{Tr} = \{\Omega_N^E \cup \Omega_N^V\}$).

2. Testing Set Ω_{NL}^{T}: For this set were selected 20 speakers, where 10 speakers are female (Ω_{NL}^{TF}) and 10 speakers are male (Ω_{NL}^{TM}). All speakers belong to IFMA bank, but are speakers who did not participate with pronunciations for the training set. In total, the testing set has 1000 male locutions and 1000 female locutions ($\Omega_{N2000}^{T} = \{\Omega_{N1000}^{TM} \cup \Omega_{N1000}^{TF}\}$).

3 Experimental Results

3.1 LVQ Training and Validation

After training the LVQ networks by all combinations of defined topology elements, it is shown in Figure 2a, Figure 2a, Figure 2b and Figure 2c, respectively, the obtained results with training of networks using the patterns C_4^{jm}, C_9^{jm} and C_{16}^{jm}.

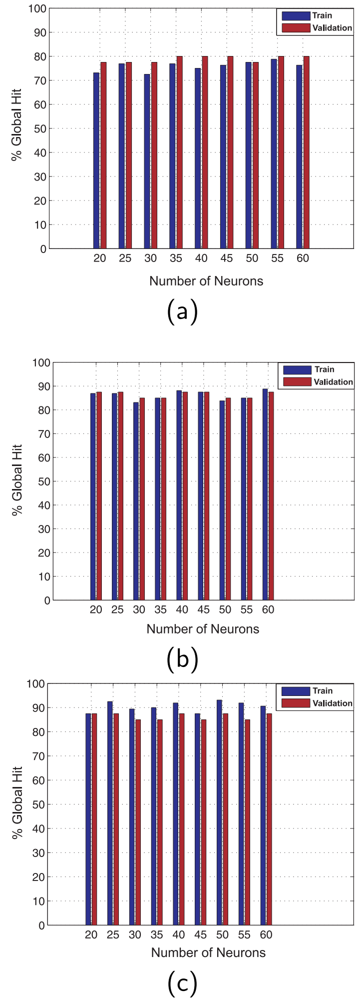

Figure 2: Result of Training and Validation Global Hit: (a) LVQ C_4^{jm}, (b) LVQ C_9^{jm} e (c) LVQ C_{16}^{jm}

3.2 LVQ Testing

The tests were applied only in trained topologies with correct classification results in the global validation greater than 80%. The obtained results (in percent) with the application of test sets Ω_{N1000}^{TM} and Ω_{N1000}^{TF} with $N = 4$, 9 e 16 are shown, respectively, in Table 3, Table 4 and Table 5.

Finishing tests, the topology with best performance was chosen according to criterion that besides presenting the highest mean results in tests must also have a reduced number of neurons.

Thus, topologies with mean hit of test detached in blue are those that showed the highest results, however topologies with mean hit detached in red have very close results but with fewer neurons. Therefore, the topology of 40, 20 and 25 neurons shown in Table 3, Table 4 and Table 5 are the topology with best result of generalization for LVQ neural network configuration.

Table 3: LVQ C_4^{jm}: Female and Male Speakers Tests

	35	40	45	55	60
Loc_F1	89	91	94	92	94
Loc_F2	95	90	90	87	88
Loc_F3	72	80	67	73	78
Loc_F4	81	84	70	67	76
Loc_F5	90	79	85	83	87
Loc_F6	98	90	95	93	89
Loc_F7	56	57	46	48	55
Loc_F8	72	76	72	64	79
Loc_F9	54	69	53	59	74
Loc_F10	68	78	79	81	77
Loc_M1	71	62	59	60	62
Loc_M2	77	77	86	85	85
Loc_M3	74	77	76	79	78
Loc_M4	72	70	79	83	79
Loc_M5	63	64	63	66	67
Loc_M6	72	65	65	65	69
Loc_M7	62	72	66	71	68
Loc_M8	70	79	74	73	75
Loc_M9	72	63	62	63	65
Loc_M10	76	81	88	78	83
Mean	74.2	75.2	73.45	73.5	76.4

Table 4: LVQ C_9^{jm}: Female and Male Speakers Tests

	20	25	30	35	40	45	50	55	60
Loc_F1	88	77	81	81	88	89	78	76	87
Loc_F2	86	88	86	85	78	87	85	85	84
Loc_F3	81	84	81	80	84	83	79	80	82
Loc_F4	87	86	70	71	81	87	69	69	88
Loc_F5	82	82	80	80	73	87	80	84	83
Loc_F6	90	83	80	80	89	90	78	79	84
Loc_F7	72	67	69	72	78	75	70	71	70
Loc_F8	87	83	79	70	90	86	72	78	82
Loc_F9	73	72	70	69	77	85	68	72	78
Loc_F10	64	68	66	62	66	64	59	64	64
Loc_M1	69	64	63	59	69	68	59	66	70
Loc_M2	81	81	76	79	78	82	73	76	82
Loc_M3	80	81	72	75	78	79	65	72	75
Loc_M4	80	69	72	71	71	80	72	75	73
Loc_M5	64	60	69	66	58	61	68	64	69
Loc_M6	90	81	72	75	86	86	74	77	83
Loc_M7	82	74	76	75	83	75	75	74	84
Loc_M8	81	69	73	72	80	71	71	69	81
Loc_M9	69	67	66	69	55	73	64	66	71
Loc_M10	77	73	76	78	73	77	73	76	76
Mean	79.15	75.45	73.85	73.45	76.75	79.25	71.6	73.65	78.3

Table 5: LVQ C_{16}^{jm}: Female and Male Speakers Tests

	20	25	30	35	40	45	50	55	60
Loc_F1	88	94	89	89	95	81	91	92	90
Loc_F2	86	95	90	88	95	79	94	88	87
Loc_F3	90	90	83	83	90	79	89	84	87
Loc_F4	89	85	87	87	84	84	88	87	88
Loc_F5	85	96	87	89	97	78	87	89	88
Loc_F6	90	99	89	90	99	82	90	89	90
Loc_F7	92	73	80	80	74	77	83	78	82
Loc_F8	98	95	97	99	93	88	97	95	97
Loc_F9	88	80	84	86	80	80	81	87	85
Loc_F10	75	80	70	73	80	76	72	73	71
Loc_M1	69	80	77	80	78	72	75	79	76
Loc_M2	84	87	81	83	88	79	87	89	84
Loc_M3	76	85	77	76	86	76	84	85	76
Loc_M4	87	88	79	83	89	81	89	87	84
Loc_M5	61	74	72	67	79	54	66	75	68
Loc_M6	78	78	80	83	86	73	87	82	78
Loc_M7	79	84	75	82	82	73	85	81	79
Loc_M8	80	84	73	82	82	73	82	77	77
Loc_M9	68	79	68	69	79	60	77	73	68
Loc_M10	76	90	80	78	88	75	77	82	77
Mean	81.95	85.8	80.9	82.35	86.2	76	84.05	83.6	81.6

3.3 *MLP Training and Validation*

In Figure 3, Figure 4 and Figure 5 are shown, respectively, the global hit results of training and validation obtained to MLP networks with one hidden layer, trained by *LM* and *RP* training algorithms using the patterns C_4^{jm}, C_9^{jm} and C_{16}^{jm}.

From the results shown in Figure 3, Figure 4 and Figure 5, it is possible to verify the influence of the set of initial weights for the MLP network. It is observed that for the same topology, one set of weights initialized randomly in a training led to satisfactory response, but in another training, the set of initial weights resulted in an undesirable response.

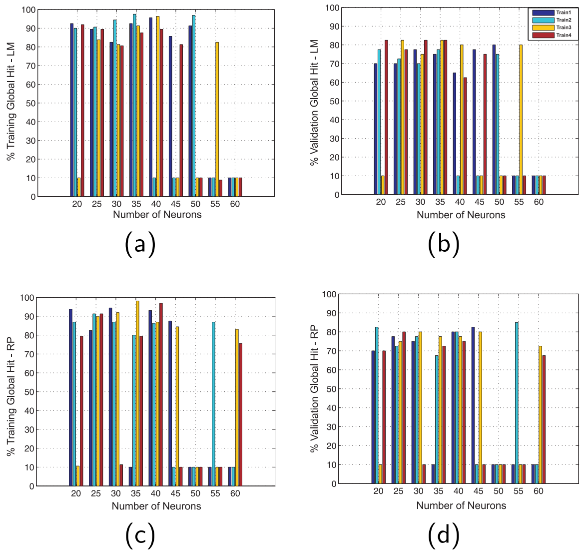

Figure 3: MLP C_4^{jm}: Result of Training and Validation Global Hit - 1 hidden layer LM and RP algorithm

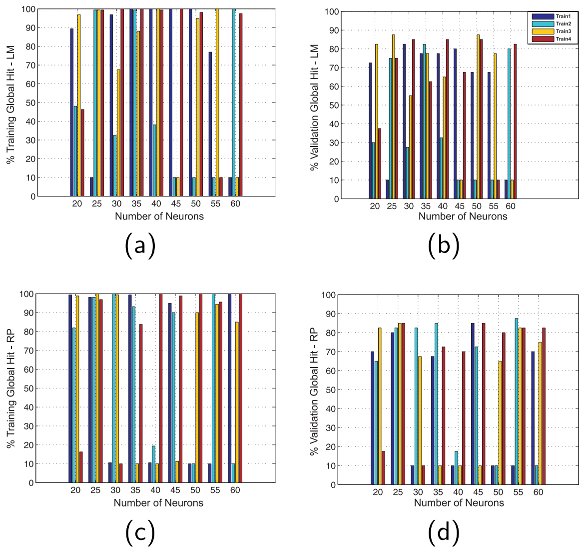

Figure 4: MLP C_9^{jm}: Result of Training and Validation Global Hit - 1 hidden layer LM and RP algorithm

3.4 *MLP Testing*

As well as carried out for the LVQ neural network, after completion of training and validation stage,

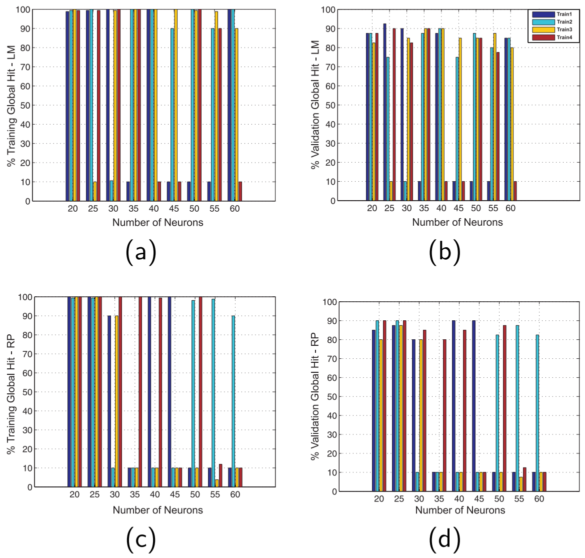

Figure 5: MLP C_{16}^{jm}: Result of Training and Validation Global Hit - 1 hidden layer for the LM and RP algorithm

between the obtained results in the simulations, the topologies of MLP Networks that showed global hit results of validation greater than 80% were selected for application testing. The best results (in percent) found in executed tests, considering the networks trained with one and two layer by RP and LM algorithms, using the test sets Ω_{N1000}^{TM} and Ω_{N1000}^{TF} with $N = 4$, 9 e 16 are shown, respectively, in Table 6, Table 7 and Table 8.

Table 6: Selection of Best Test Results MLP C_4^{jm}

Column 1	Column 2	Column 3	Column 4	Column 5
	1 layer		2 layers	
	LM	RP	LM	RP
	30-T4	45-T3	45-T1	45-T3
Loc_F1	94	82	88	89
Loc_F2	95	97	90	94
Loc_F3	83	84	82	84
Loc_F4	87	79	82	83
Loc_F5	91	88	88	88
Loc_F6	95	97	94	99
Loc_F7	67	57	65	63
Loc_F8	91	91	85	90
Loc_F9	75	59	73	70
Loc_F10	84	67	77	81
Loc_M1	72	77	77	74
Loc_M2	89	81	81	85
Loc_M3	86	82	86	85
Loc_M4	82	80	78	80
Loc_M5	72	72	71	68
Loc_M6	67	64	66	64
Loc_M7	73	68	73	75
Loc_M8	80	74	76	80
Loc_M9	73	73	70	70
Loc_M10	87	91	93	93
Mean	82.15	78.15	79.75	80.75

Finishing tests, as carried out for the LVQ networks, the topology of MLP network with best performance was chosen according to criterion that besides presenting the highest mean results in tests must also to have less complexity in the topological structure. Thus, topologies with mean hit of test detached in blue are those that showed the highest results, however topologies with mean hit

Table 7: Selection of Best Test Results MLP C_9^{jm}

Column 1	Column 2	Column 3	Column 4	Column 5
	1 layer		2 layers	
	LM	RP	LM	RP
	25-T3	45-T1	50-T4	60-T1
Loc_F1	90	93	92	95
Loc_F2	94	90	93	93
Loc_F3	93	93	91	92
Loc_F4	95	94	95	96
Loc_F5	91	97	96	96
Loc_F6	89	92	93	91
Loc_F7	76	87	85	85
Loc_F8	91	91	92	95
Loc_F9	82	76	77	83
Loc_F10	78	79	66	77
Loc_M1	74	76	83	79
Loc_M2	82	84	80	82
Loc_M3	89	87	85	91
Loc_M4	88	90	89	88
Loc_M5	65	65	64	65
Loc_M6	83	86	78	84
Loc_M7	69	73	77	68
Loc_M8	69	71	72	69
Loc_M9	88	82	84	85
Loc_M10	82	80	87	90
Mean	83.4	84.3	83.95	85.2

Table 8: Selection of Best Test Results MLP C_{16}^{jm}

Column 1	Column 2	Column 3	Column 4	Column 5
	1 layer		2 layers	
	LM	RP	LM	RP
	20-T1	50-T2	40-T1	30-T2
Loc_F1	95	88	88	92
Loc_F2	91	87	89	89
Loc_F3	97	98	98	97
Loc_F4	93	88	86	94
Loc_F5	90	93	93	95
Loc_F6	97	89	91	93
Loc_F7	92	89	94	93
Loc_F8	92	93	97	93
Loc_F9	82	85	82	79
Loc_F10	73	80	82	84
Loc_M1	84	74	74	86
Loc_M2	91	89	89	90
Loc_M3	90	86	86	91
Loc_M4	85	92	94	93
Loc_M5	59	64	67	69
Loc_M6	89	88	88	89
Loc_M7	74	79	83	88
Loc_M8	71	80	85	83
Loc_M9	82	82	83	82
Loc_M10	93	91	92	92
Mean	86	85.75	87.05	88.6

detached in red have very close results but with fewer neurons and only one layer. Therefore, the topology of 30, 25 and 20 neurons with only one hidden layer trained by LM algorithm shown in Table 6, Table 7 and Table 8 are the topology with best result of generalization for MLP neural network configuration. At the end the tests carried out with the MLP and LVQ networks and selected the topology of best performance for each configuration, for each of the three sets of patterns C_N^{jm}, $N = 4$, 9 and 16, the obtained results are summarized in the Table 9

Table 9: Final Performance Summary of the MLP and LVQ networks

	C_4^{jm}		C_9^{jm}		C_{16}^{jm}	
	Nº of Neurons	% Test Hit	Nº of Neurons	% Test Hit	Nº of Neurons	% Test Hit
LVQ	40	75.2	20	79.15	25	85.8
MLP	30	82.15	25	83.4	20	86

4 Comparison to other methods used in speech recognition

To check the performance of the methodology presented in this work, it was carried out the comparison of obtained results with MLP and LVQ neural networks with recognizers based on Gaussian Mixture Models (GMM) and Support Vector Machine (SVM). Thus, for comparison, the input parameters for the three recognizers were the same, that is, the elements C_{kn} for each pattern j. In Tables 10 and 11 test results are presented for female and male speakers, respectively (Batista e Silva, 2015; Beserra et al., 2015).

Table 10: Comparison of speech recognition results using neural networks, SVM e GMM for female speakers. OM=Order of Matrix

		OM = 2	**OM = 3**	**OM = 4**
Loc_F1	LVQ	91	88	94
	MLP	94	90	95
	SVM-Poli	68	62	65
	SVM-RBF	74	76	78
	GMM	92	84	88
Loc_F2	LVQ	90	86	95
	MLP	95	94	91
	SVM-Poli	65	65	66
	SVM-RBF	80	80	80
	GMM	94	89	82
Loc_F3	LVQ	80	81	90
	MLP	83	93	97
	SVM-Poli	60	60	77
	SVM-RBF	78	78	80
	GMM	88	88	95
Loc_F4	LVQ	84	82	96
	MLP	91	91	90
	SVM-Poli	66	72	72
	SVM-RBF	78	72	82
	GMM	70	72	83
Loc_F5	LVQ	79	90	99
	MLP	95	89	97
	SVM-Poli	66	72	72
	SVM-RBF	78	72	82
	GMM	70	72	83

Table 11: Comparison of speech recognition results using neural networks, SVM e GMM for male speakers. OM=Order of Matrix

		OM = 2	**OM = 3**	**OM = 4**
Loc_M1	LVQ	77	81	87
	MLP	89	82	91
	SVM-Poli	70	66	73
	SVM-RBF	76	80	80
	GMM	57	64	72
Loc_M2	LVQ	77	80	85
	MLP	86	89	90
	SVM-Poli	67	63	71
	SVM-RBF	76	63	81
	GMM	80	87	91
Loc_M3	LVQ	77	80	88
	MLP	82	88	85
	SVM-Poli	62	63	70
	SVM-RBF	78	80	78
	GMM	52	67	77
Loc_M4	LVQ	77	64	74
	MLP	72	65	59
	SVM-Poli	68	63	69
	SVM-RBF	76	80	80
	GMM	66	70	71
Loc_M5	LVQ	79	77	90
	MLP	87	82	93
	SVM-Poli	66	66	74
	SVM-RBF	76	80	82
	GMM	72	74	86

5 Conclusion

According obtained results, it is concluded that the MLP and LVQ configurations were able to extract the features of two-dimensional time DCT matricies of low order provided as patterns of digits to be classified. An important point verified in this paper is the influence of the values of initial weights in the result achieved by MLP Neural Network. Comparison of results for the speech recognition system using MLP and LVQ Neural Networks with other methodologies, such as SVM-polynomial, the SVM-RBF and the GMM, showed that the proposed methodology in this paper has good performance in solving of the problem in question, becoming feasible to use in speech recognition systems.

Acknowledgment

The authors would like to thank the Federal Institute of Maranhão (IFMA) and Program in Electrical Engineering of Federal University of Maranhão (PPGEE/UFMA).

References

A.Cubukcu, Kuncan, M., Kaplan, K. e Ertunc, H. M. (2015). Development of a voice-controlled home automation using zigbee module, *Signal Processing and Communications Applications Conference (SIU), 2015 23th*, pp. 1801–1804.

Batista, G. B. e Silva, W. L. S. (2015). Application of support vector machines to recognize speech patterns of numeric digits, *Natural Computation (ICNC), 2015 11th International Conference on*, pp. 831–836.

Beserra, A. A., Silva, W. L. e Serra, G. (2015). A gmm/cpso speech recognition system, *Industrial Electronics (ISIE), 2015 IEEE 24th International Symposium on*, pp. 26–31.

Cardoso, S. A., Castanho, J. E. C., Franchin, M. N. e Fontes, I. R. (2010). Sesame: sistema de reconhecimento de comandos de voz utilizando pds e rna, *Anais do XVIII Congresso Brasileiro de Automática*, Mato Grosso, Brazil, pp. 1316–1323.

Haykin, S. (2001). *Redes Neurais*, BOOKMAN COMPANHIA ED.

Silva, W. e Serra, G. (2014). Intelligent genetic fuzzy inference system for speech recognition: An approach from low order feature based on discrete cosine transform, *Journal of Control, Automation and Electrical Systems* **25**(6): 689–698.