Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 69–73
ISBN 978-83-60810-51-4

Automatic Noise Recognition Based on Neural Network Using LPC and MFCC Feature Parameters

Reza Haghmaram
Department of Electrical Engineering, IHU, Tehran, Iran
Email: haghmaram@yahoo.com

Ali Aroudi
Department of Electrical Engineering, IHU, Tehran, Iran
Email: ali.aroudi@gmail.com

Mohammad Hossein Ghezel Aiagh
Department of Electrical Engineering, IHU, Tehran, Iran
Email: ghezel@gmail.com

Hadi Veisi
Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
Email: veisi@ce.sharif.edu

Abstract—**This paper studies the automatic noise recognition problem based on RBF and MLP neural networks classifiers using linear predictive and Mel-frequency cepstral coefficients (LPC and MFCC). We first briefly review the architecture of each network as automatic noise recognition (ANR) approach, then, compare them to each other and investigate factors and criteria that influence final recognition performance. The proposed networks are evaluated on 15 stationary and non-stationary types of noises with frame length of 20 ms in term of correct classification rate. The results demonstrate that the MLP network using LPCs is a precise ANR with accuracy rate of 99.9% , while the RBF network with MFCCs coefficients goes afterward with 99.0% of accuracy.**

INTRODUCTION

IN any field of noise-corrupted speech processing such as recognition, enhancement, human machine interface, active noise control, environmental sound recognition and hearing aid, algorithms and formulas are obtained under some assumptions of noise and speech characteristics. Typically, the algorithms are designed or learned for a definite kind of noise to acquire best possible performance and adaptation in the same noisy condition. Thus for several noisy conditions, we need algorithms that each of them was trained with the present noise on each noisy environments. In designing of many speech processing algorithms, it is assumed the right noise type selection exists and they are unable to recognize the right one. As a result, in practical applications, it is necessary to recognize the right noise type to choose the appropriate model and operate correctly. This operation can be performed through an automatic noise recognition (ANR) system.

Owing to this fact that ANR is kind of classification problems, the performance given by it depends on the feature extraction method, classification method and quality/quantity of training data. Various related researches are exited which some of them investigate appropriate feature extraction methods and the other ones study various classification approaches to obtain higher performance. For example, [1] proposes Mel-frequency cepstral coefficient (MFCC), spectral flux, spectral roll-off and linear predictive coefficient (LPC), [2] suggests autocorrelation function and [3] recommends audio spectrum centroid and audio spectrum flatness as feature parameters. Moreover, For classification approach, it can be referred to hidden Markov model [4], [5], support vector machine [6], fuzzy logic system [7], [2], artificial neural network (ANN) [9], [10] and nearest neighbor method [3].

The final goal of these researches is to design an ANR algorithm having higher performance, however, it is preferred an ANR approach with minimum complexity in running mode since most speech processing algorithms have high complexity on their own. In this regard, ANN seems to be a good choice due to their architecture that consists interconnected parallel simple processing units called neurons. They also exhibit high classification performance in different applications [10], [11]. As a result, it motivates us to investigate the efficiency of ANNs such as radial basis function (RBF) and multi layered perceptron (MLP) networks in ANR application. Although numbers of researches like [9], [10] have been focused on this field, different factors and criteria such as training method, initial condition, number of neurons, feature parameters, length of frame and type and number of noises can change the evaluation performance. Thus we here explore application of MLP and RBF networks in ANR with different criteria and conditions comparing to other researches, for example using validation criterion as a factor for model generalization, LPCs and MFCCs as feature parameters, vector normalization and weights initialization, three cases for number of neurons, frames with length of 20 ms and 15 stationary and non stationary types of noises. Finally, the experimental results given by ANN classifiers are analyzed, compared to each other and the best proposed ANR is presented.

The reminder of the paper is organized as follow. In Section 2, we briefly review the architecture of RBF and MLP networks and compare them. Section 3 includes the experimental setup for evaluations such as criteria and conditions. In Section 4, evaluation results of the MLP and RBF proposed networks are presented. Section 5 gives the summary and conclusions.

STRUCTURE ARTIFICIAL NEURAL NETWORKS

Discovering an acceptable mapping between observations and classes is the final goal in classification problems. However, the mapping is unknown and can be linear or nonlinear. Owing to this fact that one of powerful applications of artificial neural networks is general function approximation, they can be used as classifiers that can approximate unknown mappings and functions. Thus we here review the MLP and RBF neural networks as a powerful tool in classification and ANR problems and compare them.

A. Multi Layer Perceptron Neural Network

The MLP neural network [11] is a fully connected feed-forward network that has one or more hidden layers followed by an output layer. Neurons of each layer possess linear activation function in their inputs and a kind of transfer function in their output as equations (1) and (2), respectively. In these equations, X, w_i, y, $g(y)$ denote input vector with components of x_i, weight of i^{th} neuron, the result of linear activation function and output of each neuron, respectively. Fig. 1 shows a neuron with activation function and transfer function. Thus each layer is a combination of neurons which calculates the weighted sum of their inputs and passes the result through a transfer function. The neurons of hidden layers contain nonlinear transfer function in their outputs as shown in Eq. (2), while neurons of output layer contain linear transfer function as Eq. (3). The hidden layers with a nonlinear transfer function give nonlinearity learning ability and nonlinear mapping to the networks beside the linear property. Whereas the network is employed as a classifier here, one additional competitive layer is needed to assign the result of output layer to one class based on a criterion. We employ Euclidian distance between the result of output layer and each class as the criterion. Therefore, the input or the result of output layer is assigned to one class that has minimum Euclidian distance to it. In order to train the model, the weights of hidden and output layers are updated using gradient-based method [12]. Here, training of the network is performed using gradient descent with momentum weight and bias learning function through quasi-Newton back-propagation training function [12]. The mean square error (MSE) custom function is used, too. Due to the fact that the training procedure is iterative, then it is continued until the MSE custom function has been minimized to zero or the error of validation subset, which shows level of generalization property, has shown an increment.

$$y = \sum_{i=1}^{n} w_i . x_i \tag{1}$$

$$g(y) = \frac{2}{1+\exp(-2y)} - 1 \tag{2}$$

$$g(y) = y \tag{3}$$

B. Radial Basis Function Neural Network

The RBF neural network [11] is a fully connected feed-forward network comprising hidden and output layers. The hidden layer is a combination of M radial basis functions with centers C_j, $j = 1,...,M$ and covariance matrixes Σ_j, $j = 1,...,M$. Each radial basis functions calculates Mahalanobis distance between input vector X and its center C_j as in Eq. (4) and passes the result through a transfer function as Eq. (5). Thus, the activation functions of the hidden layer are nonlinear. The output layer is a combination of linear neurons which calculates the weighted sum of hidden layer neurons as Eq. (1) and passes the result through a linear transfer function as Eq. (3). Whereas the network is employed as a classifier here, similar to the MLP network, one additional competitive layer is needed to assign the result of output layer to one class based on a criterion. We employ Euclidian distance between the result of output layer and each class as the criterion. Therefore, the input or the result of output layer is assigned to one class that has minimum Euclidian distance to it. In order to train the model, the centers and covariance matrices are first obtained using a random method, clustering method [13], learning-based method [14] or other methods [14], then the weights of output layer are updated based on gradient-based method. In this paper, the centers and covariance matrices are determined using k-means clustering algorithm and training of the network is done through gradient descent with momentum weight and bias learning function with the aim of quasi-Newton back-propagation training function. The MSE custom function is used, too. Due to the fact that the training procedure is iterative, then it is continued until the MSE custom function has been minimized to zero or the error of validation subset, which shows level of generalization property, has shown an increment.

$$y = \left(X - C_j\right)^t \Sigma_j^{-1} \left(X - C_j\right) \tag{4}$$

$$g(y) = \exp\left(\frac{-y^2}{2}\right) \tag{5}$$

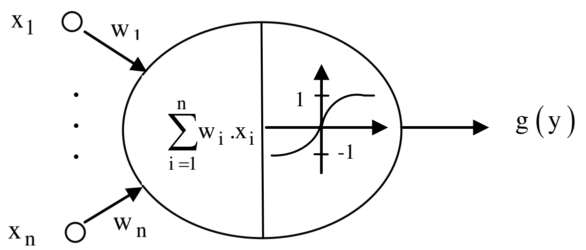

Fig. 1
MODEL OF MACCULLOCH-PITTS NEURON [11]

C. Differences Between MLP and RBF Networks

The obvious difference between MLP and RBF neural networks is the activation function of the first layers' neurons. In RBF, the activation functions of the first layer operate locally while in MLP, they operate globally. This affects the generalization properties and convergence rate of the networks. As a result, the RBF network converges faster than MLP network [16], [14]. Moreover, RBF network is less sensitive to noise owing to this fact that the locality operation of activation function does not spread the whole space of noise to the next layer. Due to these facts, it is shown that RBF network have the best approximation property while MLP network does not [17], [18]. In contrast, MLP network exhibits more generalization property, however RBF network can obtain this using more neurons in the first layer [19], [20].

EXPERIMENTAL SETUP

Here, the RBF and MLP networks are considered with one hidden, one output and one additional competitive layer. The hidden layer is chosen with $M = 12, 16$ or 30 neurons. The numbers of the output layer's neurons are equal to 15 that is the number of classes. Whereas the training of RBF or MLP does not guarantee a global optimum, initialization conditions and criteria are important to achieve better local optimum. Therefore input vectors are normalized with range of $[-1,1]$ and weights are initialized through this range to cause the training of network starts through linear region of the nonlinear transfer functions. This causes the training of neurons begins from linear region into nonlinear region, thus the final nonlinear mapping of network can be achieved better. However, if linear mapping is the final goal, it can be achieved better with stable values, too. The weights initialization is not important seriously for RBF as MLP, because the RBF network has linear transfer function, but it cause to train weights with stable values.

EXPERIMENTAL RESULTS AND EVALUATION

We have considered 15 classes corresponding to different noises types such as babble, f16, machinegun, office, volvo, white, buccaneer 1, buccaneer 2, m109, destroyerengine, destroyerops, factory1, hfchannel, leopard and pink. The noises types except for office, are taken from Noisex-92. Duration of each is about 4 minutes with sampling rate of 16kHz. We allocate 70% of length of each noise type for training procedure and the remains for test procedure. The set of training vectors is divided into two subsets with ratio of 0.8 and 0.2, respectively. The first subset is used for training networks. The second subset is the validation set which is used to prevent network from generalization property loss. Due to the fact that most speech processing methods use vectors with length of 10 to 35 ms, the experiments are performed using frames with length of 20 ms. Feature vectors like MFCCs and LPCs are considered with length of 12 coefficients, too. The MFCCs are obtained through 23 Mel-frequency filter banks.

The evaluations results for RBF and MLP classifier are presented in Table 1 in terms of accuracy rate (ACR) which is the correct classification rate for each class. It can be observed that highest accuracy rate 99.0% for the RBF network is achieved with MFCCs and $M = 30$. ACR value is obtained through average of diagonal components of confusion matrix. Table 2 demonstrates the related confusion matrix which each diagonal component shows the correct classification rate for each class and each row represents incorrect and correct classification rate for one class. The MLP network has the best accuracy rate 99.9% with LPCs and $M = 16, 30$. Table 3 shows the confusion matrix of this experiment for $M = 30$. In general, the RBF network with MFCCs has higher ACR than with LPCs for different M while this fact is inverse for the MLP network. Thus in term of ACR, MLP with LPCs outperforms the other classifiers. It proves that MLP network has more generalization properties.

Fig. 2 and Fig. 3 indicate the iteration number (ITN) and mean squared error (MSE) for each classifiers, respectively. The RBF network is noticeably low in ITN terms in contrast with MLP network. As a result, the RBF network possesses higher convergence rate than MLP network. The MSE values of MLP networks achieve minimum in comparison with RBF networks.

TABLE I.
COMPARATIVE PERFORMANCE IN TERM OF ACCURACY RATE FOR MLP AND RBF NEURAL NETWORKS WITH LPC AND MFCC FEATURE PARAMETERS

Network	M	Feature	A C R	Network	M	Feature	A C R
RBF	12	LPC	72.3%	MLP	12	LPC	93.2%
RBF	16	LPC	90.6%	MLP	16	LPC	99.9%
RBF	30	LPC	97.5%	MLP	30	LPC	99.9%
RBF	12	MFCC	81.1%	MLP	12	MFCC	86.7%
RBF	16	MFCC	91.7%	MLP	16	MFCC	98.2%
RBF	30	MFCC	99.0%	MLP	30	MFCC	98.3%

TABLE II.
CONFUSION MATRIX OF RBF NEURAL NETWORK WITH MFCCS AND 30 NEURONS IN FIRST LAYER

	Babble	F16	Machinegun	Office	Volvo	White	Buccaneer 1	Buccaneer 2	Destroyerengine	Destroyerops	Factory 1	Hfchannel	Leopard	M109	Pink
Babble	0.9531	0	0	0	0	0	0	0	0.0005	0	0.0438	0	0	0.0013	0.0010
F16	0	0.9997	0	0	0	0	0	0	0	0	0	0	0	0.0002	0
Machinegun	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Office	0	0	0	0.9997	0	0.0002	0	0	0	0	0	0	0	0	0
Volvo	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
White	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Buccaneer 1	0	0.0031	0	0	0	0	0.9968	0	0	0	0	0	0	0	0
Buccaneer 2	0	0	0	0	0	0	0	0.9867	0	0	0	0	0	0	0.0132
Destryerengine	0.0007	0	0	0	0	0	0	0	0.9989	0	0	0.0002	0	0	0
Destroyops	0	0.0002	0	0	0.0002	0	0.0005	0	0	0.9989	0	0	0	0	0
Factory 1	0.0210	0.0005	0	0	0	0	0	0.0002	0.0005	0	0.9271	0	0	0	0.0505
Hfchannel	0	0	0	0	0	0	0	0	0.0029	0	0	0.9970	0	0	0
Leopard	0	0	0	0.0002	0	0	0	0	0	0	0	0	0.9997	0	0
M109	0.0002	0	0	0	0	0	0	0	0	0	0.0005	0	0	0.9989	0.0002
Pink	0	0	0	0	0	0	0	0.0010	0	0	0.0026	0	0	0.0002	0.9960

TABLE III.
CONFUSION MATRIX OF MLP NEURAL NETWORK WITH MFCCS AND 30 NEURONS IN FIRST LAYER

	Babble	F16	Machinegun	Office	Volvo	White	Buccaneer 1	Buccaneer 2	Destroyerengine	Destroyerops	Factory 1	Hfchannel	Leopard	M109	Pink
Babble	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F16	0	0.9997	0	0	0	0	0	0	0	0	0	0	0	0.0002	0
Machinegun	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Office	0	0	0	0.9981	0	0	0.0010	0	0	0.0007	0	0	0	0	0
Volvo	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
White	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Buccaneer 1	0	0.0002	0	0	0	0	0.9992	0	0	0	0	0	0	0.0005	0
Buccaneer 2	0	0	0	0	0	0	0	0.9981	0	0	0	0	0	0	0.0018
Destryerengine	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Destroyops	0	0	0	0.0007	0	0	0.0010	0	0	0.9981	0	0	0	0	0
Factory 1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Hfchannel	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Leopard	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
M109	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Pink	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

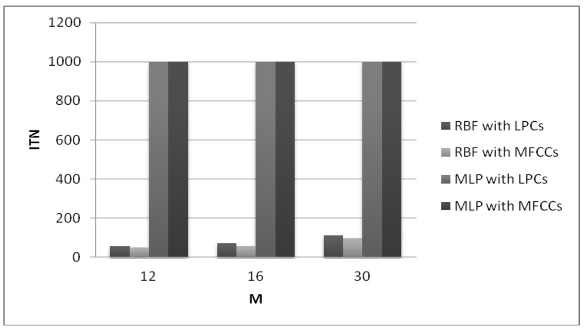

Fig. 2
ITERATION NUMBER FOR EACH CLASSIFIER

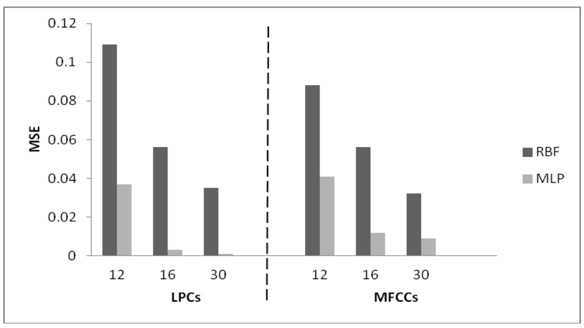

Fig. 3
Mean Squared Error Value FOR EACH CLASSIFIER

Consequently, RBF network converges fast but with high mean squared error, because the inclination rate of the generalization property reduction is very high by comparison with the MSE reduction. This fact causes the network to hold generalization but with high MSE value. Then, when low MSE and ITN is simultaneously achieved, higher ACR is accessible. However, there are a trade-off between them and they cannot occurred at once as we can see in MLP networks that high ACR are obtained with low MSE but high ITN.

Therefore, in terms of ACR, MLP network with LPCs and $M = 16$ and 30 , in terms of ITN, RBF are the best choices. However, in running mode, the network with fewer M operates faster as MLP network with $M = 16$.

Summary And Conclusions

In this paper, the automatic noise recognition has been studied using the RBF and MLP artificial neural networks based on MFCCs and LPCs. The architectures and advantages of each network are briefly reviewed. Experimental evaluations were performed on 15 stationary and non-stationary types of noises with length frame of 20 ms in term of accuracy rate. The evaluation results of networks were analyzed and compared to each other. Most proposed ANRs show acceptable accuracy rate higher than 90% while the MLP network using LPCs with 16 or 30 neurons in first layer has achieved 99.9% of accuracy rate . Beside this, RBF network with MFCCs and 16 neurons in first layer has resulted in accuracy rate of 99.0%.

References

[1] V. *Peltonen*, J. Tuomi, A. Klapuri, J. Huopaniemi, and T. Sorsa, "Computational auditory scene recognition," in *IEEE International Conference on Audio, Speech and Signal Processing,* Orlando, Florida, vol 2, pp. 1941-1944, 2002.

[2] J. Yang, and B.D. Barkana, "The acoustic properties of different noise sources," in *Proceedings of the 6th International Conference on Information Technology*, pp. 1265, 2009.

[3] J. Wang, J. Wang, K. He, and C. Hsu, "Environmental *sound classification using hybrid.* svm/knn classifier and mpeg-7 audio low-level descriptor*,"* in *Proc. Int. Joint Conf. Neural Networks, Canada,* pp. 1731–1735, July 2006.

[4] L. Ma , D. Smith, and B. "Milner, "Environmental noise classification for context-aware applications," *Proc. DEXA,* pp. 350-370, 2003.

[5] P. Gaunard, C. Mubikangiey, C. Couvreur, and V. Fontaine, "Automatic classification of environmental noise events by hidden markov models," *IEEE International Conference,* Seattle, WA, pp. 3609-3612, May 1998.

[6] B.D. Barkana, B. Uzkent, "Environmental noise classifier using a new set of feature parameters based on pitch range," *Applied Acoustics,* pp. 841-848, 2011.

[7] F. Beritelli, S. Casale, G. Ruggeri, "New results in fuzzy pattern classification of background noise," in *International Conference on Signal Processing,* pp. 11–20, Aug. 2000.

[8] B. Uzkent, Barkana B.D., Yang J., "Automatic environmental noise source classification model using fuzzy logic," *Expert Systems with Applications*, vol. 38, pp. 8751-8755, July 2011.

[9] D. Buket, Barkana, Inci Saricicek, "Environmental Noise Source Classification Using Neural Networks," in *Eventh International Conference on Information Technolog ,* pp. 259-263, 2010.

[10] F. Beritelli, R. Grasso, "A pattern recognition system for environmental sound classification based on MFCCs and neural networks," Signal Processing and Communication Systems, ICSPCS, pp. 1-4, Dec. 2008.

[11] Christopher M. Bishop, *Pattern Recognition and Machine Learning.* Springer-Verlag New York, 2006.

[12] P.E Gill, W. Murray, and M.H. Wright, *Practical optimization*. Academic Press, 1981.

[13] J. Moody, and C.J. Darken, "Fast learning in networks of locally tuned processing units," *Neural Computation,* vol. 6, pp. 281-294, 1989.

[14] S. Theodoridis, and K. Koutroumbas, *Pattern Recognition*. Elsevier Academic Press, 4th edition, 2008, pp. 136-137.

[15] D.R. Hush, and B.G. Home, "Progress in supervised neural networks," *IEEE Signal Processing Magazine,* vol. 10, pp. 8-39, Jan. 1993.

[16] S.H. Lane, M.G. Flax, D.A. Handelman, and J.J. Gelfand, "Multilayer perceptrons with B-spline receptive field functions," in *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann, San Mateo, CA, pp. 684-692, 1991.

[17] F. Girosi, and T. Poggio, "Networks and the best approximation property," *Biological Cybernetics,* vol. 63, p.p. 169-176, 1990.

[18] Christopher M. Bishop, *Neural networks for pattern recognition.*, Oxford University Press, 1995, pp. 169.

[19] E.J Hartman, J.D. Keeler, and J.M. Kowalski, "Layered neural networks with Gaussian hidden units as universal approximations," *Neural Computations,*vol. 2, pp. 210-215, 1990.

[20] S. Theodoridis, and K. Koutroumbas, *Pattern Recognition*. Elsevier Academic Press, 4th edition, 2008, pp. 134.