Comparison of Methods of Noise Classification

Antonio Nascimento[1], Felipe de S. Farias[1], Marilia Alves [2]

[1]Programa de Pos-Graduacao em Engenharia de Defesa
Instituto Militar de Engenharia (IME)

[2]Programa de Pos-Graduacao em Engenharia Eletrica
Instituto Militar de Engenharia (IME)

3

{antonio.nascimento,felipe.farias,marilia.alves}@ime.eb.br

Abstract. *Ok here we make the abstract. Needless to say it comes last. Should have max 10 lines.*

1. Introduction

Many Acoustic Signal Processing (ASP) tasks are performed in noisy conditions. The presence of noise, additive or otherwise, decreases the performance of those tasks, be it speaker recognition [Ming et al. 2007], emotion recognition [Schuller et al. 2010], source localization [Benesty 2000] or speech recognition [Friesen et al. 2001]. Thus, it is imperative to study noise so we can better assess how it affects ASP tasks and how we can deal with it. Automatic classification of types of noise is an important part of this study, since the knowledge of the kind of noise present in a given situation is useful knowledge to better treat it [May et al. 2012].

There is extensive previous work in the field of audio classification. There is great variety in the methods used to perform this task, such as statistical methods [Dal Degan and Prati 1988, Peltonen et al. 2002], methods using stochastical knowledge such as *Hidden Markov Models* (HMM) [Ma et al. 2003], using neural networks [Beritelli et al. 2005] and support vector machines [Cumani and Laface 2012]. There is variety also in the applications sought, such as speaker recognition [Kinnunen and Li 2010, Murty and Yegnanarayana 2006, Farrell et al. 1994], acoustic scene recognition [Piczak 2015, Barchiesi et al. 2015] or animal species recognition [Somervuo et al. 2006, Lee et al. 2008]. The noise classification is different only in application, but can be performed using any method used for audio classification [Beritelli et al. 2007, Ma et al. 2006].

This work proposes to evaluate the performance of four common methods used in audio classification in the specific task of classify noise. To this end, we implement those methods in the same set of audio files containing different types of noise. These files are taken from the NOISEX database [Varga and Steeneken 1993], a database comprised of audios of 15 types of noise. The evaluation follows these steps: The extraction of attributes of each audio file, construction of the models according to each method, classification and evaluation of the results. The methods compared are the Neural Network, Gaussian Mixture Model, Support Vector Machines and K-means.

The remainder of this paper is organized as follows. Section 2 introduces the task of noise classification, as long as the methods used in this paper. Section 3 describes the experiments performed and the results obtained and, finally, in Section 4 we present our conclusions about the results found, as long as the future works.

2. Noise Classification

2.1. Extraction of Audio Attributes

The first step towards classification is the extraction of attributes from the data that are useful for the classification algorithms. In this Section we present two of the most widely used in the literature, the Linear Predictive Coefficient [Rabiner and Juang 1993] and the Mel-Frequency Cepstral Coefficient (MFCC) [Xu et al. 2005]. In this work, we will use the MFCC to represent our audios.

2.1.1. Linear Predictive Coefficient (LPC)

In the Linear Predictive analysis of audio, the audio is divided into frames of the same size, usually 20ms, and each frame is predicted as the linear weighted sum of the *n* previous frames, there *n* represents the order of the prediction [Rabiner and Juang 1993].

$$\hat{s} = \sum_{k=0}^{n} \alpha_k s(n-k) \tag{1}$$

The difference between the prediction and the actual values of the frame is computed as error. The coefficients α_k are obtained minimizing the prediction error through the least squares minimization.

2.1.2. Mel-Frequency Cepstral Coefficient (MFCC)

The Mel-Frequency cepstrum is efficient in modeling pitch and frequency content of audio signals. It yields better results when coding audio for in classification tasks than the LPC [Li et al. 2001].

In the mel-cepstral analysis, the audio signal is filtered by *K* bandpass filters, which have constant mel-frequency interval and cover the $0 - 4000$Hz frequency range. The MFCCs are calculated by the following equation:

$$c_n = \sqrt{\frac{2}{K}} \sum_{k=1}^{K} (\log S_k) \cos[n(k-0.5)\pi/K] \tag{2}$$

In which c_n is the coefficient for the n^{th} frame and $S_k, k = (1, 2, ..., K)$ are the output of each bandpass filter.

In this work, we used $K = 13$, and the $\delta_n = c_n - c_{n-1}$ and $\delta\delta = \delta_n - \delta_{n-1}$ coefficients were calculated, thus rendering 39 coefficients by frame.

2.2. Classification Methods

2.2.1. K-means

This is a method used in summarization. Didn't find references.

2.2.2. Gaussian Mixture Models (GMM)

The GMM is used based on the knowledge that a set of acoustic signal classes can be represented by it's component densities [Reynolds and Rose 1995]. The GMM is a weighted sum of K component densities, given by:

$$p(\vec{x}|\lambda) = \sum_{k=1}^{K} p_k b_k(\vec{x}) \tag{3}$$

Where $\vec{x}$ is a random vector representing our input signal, K is the number of components of the signal, $p_k, k = 1, ..., K$ are the mixture weights and $b_k(\vec{x}), k = 1, ..., K$ are the component densities.

Each class is represented by it's model λ:

$$\lambda = \{p_k, \vec{\mu}_k, \sum k\}, k = 1, ..., K \tag{4}$$

Where $\vec{\mu}_k$ is the mean and $\sum k$ is the covariance matrix of each component density $b_k(\vec{x})$.

2.2.3. Support Vector Machines (SVM)

The SVM is a machine learning technique that has successfully been used in pattern recognition tasks, such as audio classification [Dhanalakshmi et al. 2009]. The basic idea of this technique is to estimate the hyperplane that better separates a group of data [Cumani and Laface 2012]. The hyperplane can be linear or be created using a kernel function to better do the separation.

There are two distinct phases: in the training phase, the algorithm estimates the best hyperplane, searching for the one that maximizes the distance between the training data of different classes. In the test phase, the classification using the trained hyperplane is performed using different data. This way, we can assess the generalization capacity of the model.

KERNEL QUADRATICO UTILIZADO

2.2.4. Neural Network

Neural Network is an interconnected group of artificial neurons designed to simulate the functionality of a brain. It is organized in layers: the input layer, the hidden layer and the output layer. Each neuron has it's own weight and a activation function [Wu et al. 2007].

In the classification task, the weights of the neurons are updated by training the network with labeled data, until the network can yield the expected result in the output layer.

50 NEURONS UTILIZADOS

3. Experiments

3.1. Experimental Setup

The experiments were performed using MATLAB, using the Neural Network toolbox and the SVM toolbox. They were performed in two separated computers. The K-means and GMM experiments were performed with an Intel-i7 1,8GHz processor, with ram memory of 8GB and Windows10 OS. The Neural Network and SVM experiments were performed with an an Intel-i3 processor, with ram memory of 4GB and Linux OS.

3.2. Database Description

For this work we used the NOISEX database [Varga and Steeneken 1993]. This database is composed by 15 audio files, one for each of the 15 different classes of noises, which are shown in Table 1. All the audio files are in .wav format, 3 minute and 55 seconds long and they have a bit rate of 319 kbits per second.

Table 1. Classes existing in the NOISEX database.

babble	buccaneer 1	buccaneer 2	destroyer engine	destroyer operations room
f16	factory floor 1	factory floor 2	HF channel	leopard
m109	machine gun	pink noise	volvo	white noise

3.3. Cross-Validation

The 4-fold cross-validation was performed. The accuracy reported in Table 3 are the mean of the results in each test. This way, the generalization of each model is better asserted.

3.4. Results

You may be wondering why have only the 'results' and not the 'experiments' subsection. This is because we already told the reader everything he has to know in the previous sections. Section 2 describes the different methods and Sections 3.1 and 3.2 describe the details of the experiments.

4. Conclusions

Here we conclude the paper. I suck at this, so pls someone do it for me. The one thing I know is that we have to summarize our findings and link it to our problem stated in the introduction, telling the reader whether we were successful or not in the task we proposed in the beginning.

Table 2. Confusion Matrix for Neural Network

	Babble	Bucc 1	Bucc 2	Dest Eng	Dest Ops	F16	Fac 1	Fac 2	HFChn	Leop	M109	MachGun	Pink	Volvo	White
Babble	6,5%						0,1%	0,1%							
Buccaneer 1		6,6%													
Buccaneer 2			6,7%												
Destroyer Engine				6,6%											
Destroyer Ops					6,5%										
F16						6,6%									
Factory 1							6,0%	0,2%					0,1%		
Factory 2							0,3%	6,4%							
HF Channel									6,7%						
Leopard										6,6%					
M109											6,6%				
Machine Gun												6,6%			
Pink							0,2%						6,6%		
Volvo														6,6%	
White															6,7%

Table 3. Accuracy per class for the methods compared.

Class	K-means	GMM	Neural Network	SVM
Babble	88,2%	98,9%	97,5%	
Bucanneer 1	96,6%	99,1%	99,1%	
Bucanneer 2	98,8%	99,7%	99,8%	
Destroyer Engine	99,8%	99,7%	99,7%	
Destroyer Ops	90,8%	96,9%	98,3%	
F16	95,2%	99,10%	97,6%	
Factory 1	59,3%	87,6%	92,3%	
Factory 2	93,9%	95,0%	94,7%	
HF Channel	100,0%	100,0%	99,9%	
Leopard	99,1%	99,6%	99,4%	
M109	94,1%	99,3%	99,3%	
Machine Gun	7,1%	99,5%	99,5%	
Pink	99,7%	98,0%	96,9%	
Volvo	90,8%	99,4%	99,7%	
White	99,9%	99,9%	100,0%	
OVERALL	89,2%	98,0%	98,4%	

References

Barchiesi, D., Giannoulis, D., Stowell, D., and Plumbley, M. D. (2015). Acoustic scene classification: Classifying environments from the sounds they produce. *IEEE Signal Processing Magazine*, 32(3):16–34.

Benesty, J. (2000). Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *The Journal of the Acoustical Society of America*, 107(1):384–391.

Beritelli, F., Casale, S., and Serrano, S. (2005). Adaptive robust speech processing based on acoustic noise estimation and classification. In *Signal Processing and Information Technology, 2005. Proceedings of the Fifth IEEE International Symposium on*, pages 773–777. IEEE.

Beritelli, F., Casale, S., and Serrano, S. (2007). Adaptive v/uv speech detection based on acoustic noise estimation and classification. *Electronics Letters*, 43(4):249–251.

Cumani, S. and Laface, P. (2012). Analysis of large-scale svm training algorithms for language and speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(5):1585–1596.

Dal Degan, N. and Prati, C. (1988). Acoustic noise analysis and speech enhancement techniques for mobile radio applications. *Signal Processing*, 15(1):43–56.

Dhanalakshmi, P., Palanivel, S., and Ramalingam, V. (2009). Classification of audio signals using svm and rbfnn. *Expert systems with applications*, 36(3):6069–6075.

Farrell, K. R., Mammone, R. J., and Assaleh, K. T. (1994). Speaker recognition using neural networks and conventional classifiers. *IEEE Transactions on speech and audio processing*, 2(1):194–205.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, 110(2):1150–1163.

Kinnunen, T. and Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech communication*, 52(1):12–40.

Lee, C.-H., Han, C.-C., and Chuang, C.-C. (2008). Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1541–1550.

Li, D., Sethi, I. K., Dimitrova, N., and McGee, T. (2001). Classification of general audio data for content-based retrieval. *Pattern recognition letters*, 22(5):533–544.

Ma, L., Milner, B., and Smith, D. (2006). Acoustic environment classification. *ACM Transactions on Speech and Language Processing (TSLP)*, 3(2):1–22.

Ma, L., Smith, D., and Milner, B. P. (2003). Context awareness using environmental noise classification. In *Interspeech*.

May, T., Van De Par, S., and Kohlrausch, A. (2012). Noise-robust speaker recognition combining missing data techniques and universal background modeling. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):108–121.

Ming, J., Hazen, T. J., Glass, J. R., and Reynolds, D. A. (2007). Robust speaker recognition in noisy conditions. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5):1711–1723.

Murty, K. S. R. and Yegnanarayana, B. (2006). Combining evidence from residual phase and mfcc features for speaker recognition. *IEEE signal processing letters*, 13(1):52–55.

Peltonen, V., Tuomi, J., Klapuri, A., Huopaniemi, J., and Sorsa, T. (2002). Computational auditory scene recognition. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 2, pages II–1941. IEEE.

Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*, pages 1–6. IEEE.

Rabiner, L. and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Reynolds, D. A. and Rose, R. C. (1995). Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE transactions on Speech and Audio Processing*, 3(1):72–83.

Schuller, B., Vlasenko, B., Eyben, F., Wollmer, M., Stuhlsatz, A., Wendemuth, A., and Rigoll, G. (2010). Cross-corpus acoustic emotion recognition: Variances and strategies. *IEEE Transactions on Affective Computing*, 1(2):119–131.

Somervuo, P., Harma, A., and Fagerlund, S. (2006). Parametric representations of bird sounds for automatic species recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6):2252–2263.

Varga, A. and Steeneken, H. J. (1993). Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech communication*, 12(3):247–251.

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y.-X., Chang, Y.-F., and Xiang, Q.-L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pages 11–16. IEEE.

Xu, C., Maddage, N. C., and Shao, X. (2005). Automatic music classification and summarization. *IEEE transactions on speech and audio processing*, 13(3):441–450.