Audio Proccessing and Indexing
Bird sound Classification

Athanasios Agrafiotis s2029413

1 Problem

The concept of this assignment is classify unlabelled sounds of birds. The dataset contains sound of birds, noise sounds and at last sound of birds which made by humans. The task of classify those sounds is really hard considering that our data are unlabelled. In order to accomplish the assignment unsupervised learning will be used on the data.

2 Introduction

Unsupervised learning has been used widely and especially for sound classification and text. The domain of unsupervised learning is try to classify unlabeled data based on the features. One of the most important task to achieve good performance is the **Feature Engineering** which is a process of choosing the right metrics for a machine learning algorithm. The machine learning algorithms which have been used for unsupervised learning is the K-means, principal component analysis. The important aspect is to find the features which can represent the data as natural groups, clusters.

3 Related Work

The features which have been used seems to affect the result and classifier which have been used in order to classify. In recent research the mel-frequency ceptral coefficients (MFCC) which consider a an efficient and simple way [2].
Another research which has been used a baseline the of unsupervised learning is the Mel-scpectogram and the random forest classifier [1].

4 Dataset

The dataset have 56 sound files. In order to classify the sounds we had to preprocess the data and extracting the features. Different preprocess seems to effect the results. The dataset was splitted in train set 80% and test set 20% the random state was 0. In table 1 it is demonstrated the test set annotated by human. As noise it consider human conversations or very deep sound by car.

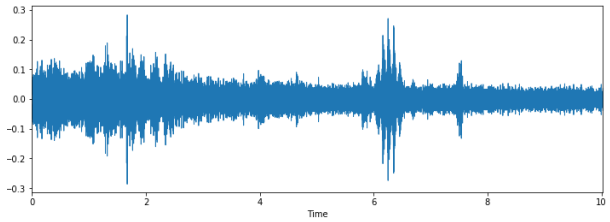

Figure 1: Noise only

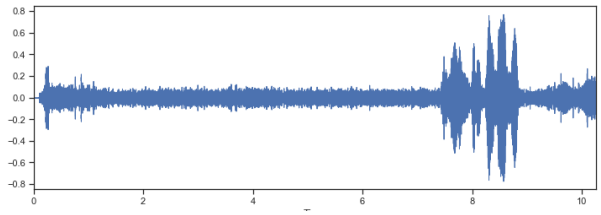

Figure 2: Bird sound and Human voice

5 Research

For the purpose of this assignment working on a sample of bird sounds. In order to classify each sound features were extracted. The test was made by extracting the mel-frequency ceptral coefficients (MFCC).
Mel-frequency ceptral coefficients: In order to extract the feature of mel-frequency ceptral coefficients. It create a short analysis window.We compute the fourier transform of the window. The result of the fourier transform it creates a spectrum . The spectrum represents the Amplitude in the frequency axis. Next we take a window which maps the power in a mel scale. Then we obtain the power of each of the mel-frequencies. And at last it calculates the discrete cosine trans-

sound	label	Knnlabel
a0484d7e	bird	0
9fb8c1db	bird	0
9f4b1924	bird	1
a0025b2	noise-bird	0
9fff79d1	noise	1
9f252aab	noise	0
a02ac7bc	noise-bird	0
9f547a54	bird	1
a0450539	bird-noise	1
9ff9291c	bird-noise	0
9fa17f89	bird-noise	0
9ff0f64c	bird	0

Table 1: Test set

form of each the mel-log powers. The corrsesondig ampitudes of the result is the MFCC. The mel-frequency analysis of speech is based on human ear. It concetrates on only certain frequency components.

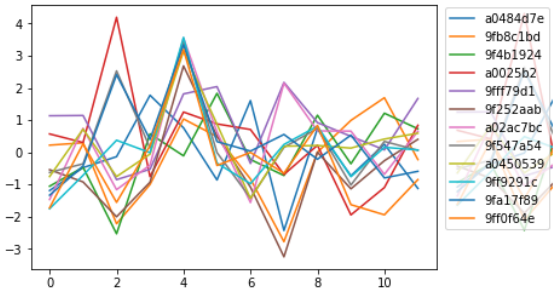

Figure 3: Mel-frequency ceptral coefficients test set

K-means clustering: In order to label the data. K-means is an unsupervised algorithm which can create clusters. The procedure is based on single steps when given data is set to a certain number of clusters. Next the number of centroids is defined known as **k**. Next the centroids are placed randomly and each of the points is associated with each of the centroids.

- Place K points into the space represented by the data
- assigned each data to the group that has the closest centroid. Based on the squared Euclidean distance. In the formula bellow it calculate the c_i the colection of centroids in a set C and then each data point x is assigned to a cluster.

$$argmin_{/} dist(c_i, x)^2 \tag{1}$$

- Nect each centroid updated by step: The centroids are recomputed. In this step the mean of all data points assigned to that centroids's cluster.

$$c_i = \frac{1}{|S_i|} \sum x_i \in S_i^{x_i} \tag{2}$$

The convergence of the algorithm after a number of iterations is guaranteed. The result is a local optimum.

6 Conclusion

The purpose of the research was to classify sounds if it was bird or only noise binary classification. The mel-frequency coefficients were used as a feature. In order to evaluate the performance of the classifier a sample was used which labelled by myself this method is known as weak supervision. The results depict that each audio which contains bird sound is 0 and each which contains noise is labelled as 1. The outcome is consider as good as 3 out of 12 sounds misclassified 75% classified correctly. Without the noisy data 3 out of 7 sounds misclassified which means that 57% classified correctly.

References

[1] Leo Breiman. "Random Forests". In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

[2] Dan Stowell et al. "Bird detection in audio: a survey and a challenge". In: *CoRR* abs/1608.03417 (2016). arXiv: 1608.03417. URL: http://arxiv.org/abs/1608.03417.