Bird Species Classification Based on Mixed-GMM

Long CHENG and Hua-qing ZHANG
Information Engineering School, Communication University of China, Beijing, China

Keywords: Mel-Frequency Cepstral Coefficients (MFCC), Mixed Gaussian Mixture Model (mixed-GMM), Call-song model (CSM), Male-female model (MFM).

Abstract. A bird species classification method is raised on the basis of Mel-Frequency Cepstral Coefficients (MFCC) and mixed Gaussian Mixture Model (mixed-GMM). The mixed-GMM is established according to the species of birds by classifying each bird recordings into call-song model (CSM) and male-female model (MFM), and extracting each species’ feature parameter respectively. In the recognition phase, each unknown bird recording’s MFCC is matching with the known CSM and MFM, and to classify it into the species which gets the highest score with its model. Results show that the accuracy rate of mixed-GMM is higher than conventional GMM.

Introduction

Unlike other animals, birds move widely and uncertainly, and most of them have seasonal migration habits. Bird habitat is proved to be very secluded, so is difficult to see them. But we are able to make the recognition by recording their chirps.

Then these recordings need to be preprocessed, after that we need a feature extraction block technique to characterize the preprocessed chirps. The next step is using the known species’ characteristic values to build parameter model. Finally we can classify the unknown birds’ recordings with these models.

In this paper, a bird species classification method is proposed and demonstrated on the basis of Mel-Frequency Cepstral Coefficients (MFCC) and mixed Gaussian Mixture Model (mixed-GMM). It can offer a higher recognition rate compared with the conventional GMM.

Classification Based on Mixed-GMM

In this section experimental setup for the classification based on mixed-GMM is described and the classification system based on mixed-GMM is shown in Figure 1.

Preprocessing

The preprocessing stage includes signal normalization, pre-emphasis filtering and removal of silence intervals.

After importing the bird recordings and changing them into input signal, the dynamic range of the signal amplitude will be mapped into the interval from 0 to +1. Then the signal passes through a high-pass pre-emphasis filter. And the silence intervals can be removed using a voice activity detection (VAD) technique.
Psychophysical studies have shown that human perception of the sound frequency contents for speech signals does not follow a linear scale. Thus for each tone with an actual frequency, \( f \), measured in Hz, a subjective pitch is measured on a scale called the ‘Mel’ scale. The relation between actual frequency and Mel frequency is provided by (1).

\[
f_{\text{mel}} = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)
\]

(1)

In order to get the MFCC, the preprocessed signal needs a series of calculation. First, the signal is divided into frames and to add the window function on them. After that each frame need to be mapped from time domain to frequency domain by using Discrete Fourier Transform (DFT). The DFT equation is defined in (2), where, \( 1 \leq k \leq N \), and \( N \) denotes N-point DFT.

\[
Y(k) = \sum_{n=1}^{N} y(n) e^{-j \frac{2 \pi n k}{N}}
\]

(2)

Next, the results of DFT pass through the Mel-scaled band-pass filters and then calculate the logarithmic energies of filters’ output.

Finally, the distributed discrete cosine transform (DCT) is applied to the log filter bank energies \( S(m) \) to de-correlate the energies and the final MFCC coefficients \( C_n \) are provided in (3).

\[
C_n = \sum_{m=0}^{M-1} S(m) \cos \left( \frac{\pi n (m-0.5)}{M} \right)
\]

(3)

Where, \( M \) denotes the number of Mel-scaled band-pass filters, and \( n=0, 1, 2, \ldots, L-1 \), and \( L \) is the desired number of MFCCs.

**Mixed-GMM Training**

GMM is a multidimensional probability density function model on the basis of probability statistic method. And GMM is a weighted sum of \( M \) Gaussian component densities. A GMM model is represented by the class model \( \lambda \) shown in (4).
\[ \lambda = \{ p_i, \mu_i, \Sigma_i \}, \ i = 1,2 \ldots M \] (4)

Where, \( p_i \) are the mixture weights, \( \mu_i \) are the mean vectors and \( \Sigma_i \) are the covariance matrices.

At the model training stage, a conventional bird species GMM model is represented by a model \( \lambda \). However, the mixed-GMM model is divided into CSM model and MFM model, which is represented by more than one \( \lambda \). The Expectation Maximization (EM) algorithm is most commonly used to iteratively derive optimal class models.

**Classification**

In this phase, we calculate the probabilities between each test data set and all models by using Maximum a posteriori (Map) method. The probabilities are referred to by the probability scores based on logarithm, so we can accomplish the classification according to the scores.

Because each mixed-GMM model is represented by more than one single \( \lambda \), we should select the highest probability score as this model’s final score.

**Experimental Results and Discussions**

**Experimental Results**

In this section, the performance of bird species classification based on different model training methods was evaluated. For the purpose of the principle of single variable, the experiment used the same birds recording data. All experimental bird recordings came from the website (http://www.xeno-canto.org/). The model training samples were collected within 7 bird species, more than 200 and each species used 20 to 40 samples depend on the number of model \( \lambda \) for the model building. And test samples were collected within the same bird species, 110 in total but differ from the training samples. Each species selected 10 samples to test.

![Figure 2. Probability scores map based on GMM](image-url)
Figure 3. Probability scores map based on mixed-GMM.

Figure 2 shows the probability scores of bird species classification based on conventional GMM training method.

While Figure 3 shows the probability scores based on mixed-GMM training method. The x-axis shows species model, and y-axis are represented by test recordings. The probability scores of test recording data are shown as blocky rectangles of color.

Results Discussion

By comparing Figure 2 with Figure 3 we can clearly see that image diagonal of the probability scores map of bird species classification based on mixed-GMM training method is more distinct than the conventional GMM one. That is because the whole probability scores enhance by increasing the GMM model number and for the correct species models is more obviously.

And the results show that the percentage of accuracy rate for the conventional GMM is 81.43% while mixed-GMM improved the performance significantly to 92.86%.

Figure 4. Comparative analysis based on different methods.

The correct and false count comparison between conventional GMM method and mixed-GMM method is shown in Figure 4.
Conclusion

In this research a new approach for the model building to improve the performance of classification system has been identified. The conventional GMM model building did not accurately capture the exact characteristics of the bird recording which contains the bird species specific information, like bird call and song, male and female. Improvement in the characteristic capture was found by using mixed-GMM.

This new approach for bird species classification is promising and may provide an improvement in results achieved in other studies.

References


