Random Forest Algorithm for Recognition of Bird Species using Audio Recordings

P. V. Lavanya Sudha[1], Dr. G. Lavanya Devi[2,] and Naresh Nelaturi[3]

[1]M.Tech, Department of Computer Science and System Engineering, Andhra University College of Engineering, Visakhapatnam, India

[2]Assistant Professor, Department of Computer Science and System Engineering, Andhra University College of Engineering, Visakhapatnam, India

[3]Research Scholar, Department of Computer Science and System Engineering, Andhra University College of Engineering, Visakhapatnam, India

[1](Email: pvslavanya95@gmail.com)

[2](Email: lavanyadevig@yahoo.co.in)

[3](Email: naresh855@gmail.com)

Abstract:

Birds are the part of the whole ecosystem. Real world audio data faces certain difficulties such as multiple birds in the same recording, other sources of non-bird sounds (e.g. buzzing insects), and background noises like wind, rain, and motor vehicles. This problem is formulated as a multi-label classification problem. The problem states that every audio file can have multiple bird sounds (instances) in it, and each instance of a bird sound can correspond to one of many different birds (labels). The proposed representation uses a 2D-supervised time-frequency segmentation which can separate bird sounds that overlap in time. By using the segment features and histogram of features, the random forest classifier is developed to predict the set of bird species which is present in a given ten-second audio recording. This method achieved an accuracy of .8396.

Keywords: *Multi-Label Classification, 2D-supervised time-frequency, Random Forest Classifier.*

1.Introduction:

Birds are the part of the whole ecosystem, the interaction between human and birds occurs in several scenarios. Birds are numerous and easier to detect than other animal species. Identification of Bird Species from the audio recordings is used in many applications such as to keep track of environmental quality and to prevent the collisions between birds and plane. Using recordings produced by birds, the identification task can be done by using signal processing techniques and machine learning algorithms. The classification of birds is a form of acoustic event classification (AEC) where the target events are bird calls and songs. The audio recordings of bird sounds tend to be noisy as they are recorded in open environments and it is common to have sounds from multiple birds or other non –bird species such as insects. So, it is difficult to label such data manually due to noise, the similarity of certain bird sounds and the potentially large number of possible bird species for a given location.

The major challenges that are going to face in this dataset are:

1. Presence of background noises such as rain, wind and motor vehicles.

2. Absence of any bird sounds in the recordings.

3. Presence of multiple bird sounds in the same recording.

4. Overlapping bird sounds.

The overall structure of this research paper is as follows. Section 2 presents related work. Proposed approach is given in Section 3. Section 4 discusses about the implementation of the proposed model and reveals the results. Finally, Section 5 gives conclusions along with future scope of the work.

2. Related Work:

F. Briggs et al. used an ensemble classifier combined with a histogram of segments for multi-label classification of bird species. This method is compared with binary relevance and three machine learning algorithms. Experiments are lead on the real world dataset and shown that this proposed method got a better result using binary relevance approach.

B.Lakshminarayanan et al. an algorithm which transformed an input audio signal into a bag-of-instances representation suitable for use with MIML classifiers and used a 2D time-frequency segmentation of the audio signal, which can separate bird sounds that overlap in time.

R.Raich et al. explained about the Multi-instance multi-label learning (MIML) framework for supervised classification where bags of instances associated with multiple labels are classified.

L. Breiman. used random selection of features to split each node yields error rates using random forest algorithm.

Herrera.F et al. introduced all the needed concepts to understand multilabel data characterization, treatment and evaluation.

3. Proposed Method:

a) Pre-Processing:

The raw audio signal is onverted into a spectrogram image using parameters window size in 512, hamming window 75% overlap by dividing it into frames and applied FFT to each frame. The frequency profile of stationary noise (such as wind and streams) is identified from low energy frames, then the spectrogram is faded to reduce the background noises while preserving bird sound.

b) Generating features from the segments:

Spectrogram: Each spectrogram is divided into a collection of region's using a supervised time-frequency segmentation algorithm. The 20 spectrograms have been taken manual examples of correct segmentation, by drawing over the areas corresponding to bird sound in red, and rain-drop in blue. Because there are a large number of pixels in each spectrogram, these are subsampled 30% of red pixels as positive examples, 30% of blue pixels as negative examples, and 4% of uncolored pixels as negative examples. From the 20 manually generated spectrograms, this sampling process yields 467,958 examples.

Each pixel is described by a feature vector with the following elements:

- The raw pixel intensity of all pixels in a 17 × 17 box around the pixel (this gives a 172 = 289-d feature).
- The average intensity of all pixels in that box (1-d).
- The y-coordinate of the pixel, which corresponds to a frequency (1-d).
- The raw pixel intensity of all pixels in the same column as the pixel (256-d).

A Random Forest classifier is trained on the positive and negative examples. Then the trained Random Forest classifier is applied to each pixel in every spectrogram, which gives a probability for the pixel to be bird sound. The probabilities may be noisy when viewing individual pixels in isolation, so they are averaged over a neighborhood by applying a Gaussian blur to an image of the probabilities, with a kernel parameter $\sigma = 3$. The blurred probabilities are then compared to a threshold of 0.4. Pixels with probabilities above the threshold are considered to be bird sound and pixels with probabilities below the threshold are considered the background.

Segment-wise features: After the segments are separated, a histogram-of-gradients feature is used to generate a feature vector for each of the segments.

Bag-of-words feature for each recording: Note that the number of segments in an image may vary, so applying a machine learning algorithm to the segment features directly will be cumbersome. Instead of doing that, we generate bags-of-features from the histograms. Bags of-features (commonly called bag-of-words) is a technique commonly used in text processing that provides information with how often a particular feature occurs in an instance. Basically, they are a way to accumulate information from different features into a single training instance. The output is a single fixed-length feature vector that can then be used for classification in a MIML problem. The final histogram-of-gradients features are 100x1 vector, and these are taken as an input to the machine learning algorithm.

c) Classification

Random forests have been widely used for multi-label classification. Random Forest is operated by constructing a decision tree structure by the training examples. One of the popular algorithms is tree bagging, in which the training process includes repeatedly selecting a bootstrap sample of the training set and fitting the trees to them. After the training process, the label decision is made either on the majority of the votes or a weighted combination from individual trees.

4.Experiments and Results:

4.1 Dataset:

The proposed methods are applied to the audio dataset which was collected in the H. J. Andrews (HJA) Long-Term Experimental Research Forest, in the Cascade mountain range of Oregon, and Oregon State University Bioacoustics group have collected over 10TB of audio data in HJA using Song meter audio recording devices. The bird's dataset is the model of the relationship between 645 ten-second audio recordings and bird species. A Song meter has two omnidirectional microphones and records audio in WAV format to flash memory. A Song meter can be left in the field for several weeks at a time before either its batteries run out, or its memory is full. In this dataset, it includes rain and wind and represented a sample from 2009 and 2010 during summer of audio recording at 13 different locations. The dataset consists of 645 ten-second audio recordings in uncompressed in WAV format and there are 19 species of bird in the dataset (Table 1). Each ten-second audio recording was paired with a set of bird species that were present. There is some relevant information in WAV files about location, date and time where every 13 different locations have a distinct location code. Because each clip was recorded in a natural setting, it may contain environmental noises such as wind or rain.

Code	Name
0	Brown Creeper
1	Pacific Wren
2	Pacific-slope Flycatcher
3	Red-breasted Nuthatch
4	Dark-eyed Junco
5	Olive-sided Flycatcher
6	Hermit Thrush
7	Chestnut-backed Chickadee
8	Varied Thrush
9	Hermit Warbler
10	Swainson's Thrush
11	Hammond's Flycatcher
12	Western Tanager
13	Black-headed Grosbeak
14	Golden Crowned Kinglet
15	Warbling Vireo
16	MacGillivray's Warbler
17	Stellar's Jay
18	Common Nighthawk

Table 1: Bird Present in the Dataset

4.2 Results:

These spectrograms (Figure 2) are computed by dividing the WAV signal into overlapping frames and applying the FFT with a Hamming window. The FFT returns complex Fourier coefficients. To enhance contrast, first, normalize the spectrogram so that the maximum coefficient magnitude is 1, then take the square root of the normalized magnitude as the pixel value for an image.

The spectrogram has time on the x-axis (from 0 to the duration of the sound), and frequency on the y-axis. The maximum frequency in the spectrogram is half the sampling frequency (16kHz/2 = 8kHz).

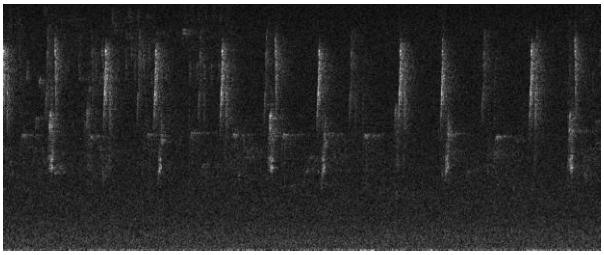

Figure 2: Spectrogram

The frequency profile of stationary noise (such as wind and streams) is estimated from low energy frames, then the spectrogram (Figure 3) is attenuated to suppress the background noise while preserving bird sound

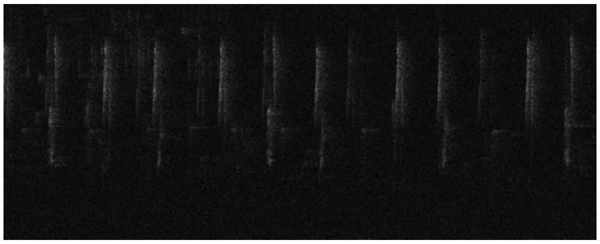

Figure 3: Noise reduced Spectrograms

These are the spectrograms (Figure 4) with the outlines of segments drawn on top of them. These segments are obtained automatically using a segmentation algorithm that is trained.

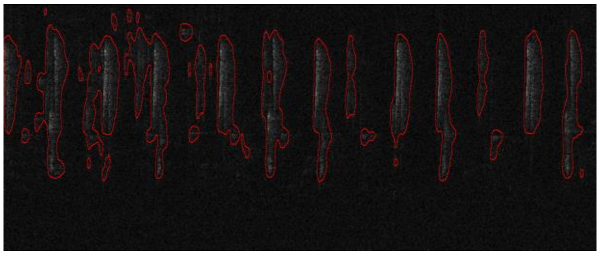

Figure 4: Segmented Spectrograms

The overall accuracy of the classification is measured by the area under the ROC(Receiver Operating Curve).The ROC is a curve that plots the true positive rate on y-axis versus the false positive rate on x-axis.

According to this model, the ROC curve (Figure 5) is generated by taking input as the labels generated by the algorithm, along with the predictions.

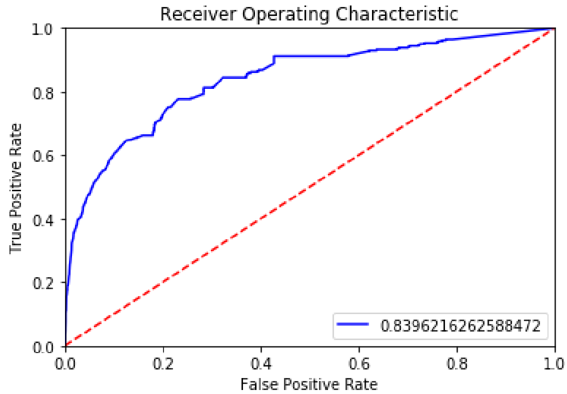

Figure 5: ROC Curve for the model

5.Conclusion:

An effort was made to understand the multi-label classification problem. The problem that attempted to solve is summarized as "Given a recording (pre-shortened to 10 s), predict if there are any birds in the recording, and if so, how many and what species (of a subset of 19 species)".

There are many challenges in going from audio to labelling of bird species, and through this project, intended to understand these challenges and gain experience using the tools needed to overcome them as move to a real-world setting. These challenges include:

1. Noise reduction
2.Separating overlapping bird sounds (the cocktail-party problem)
3.Localising birds sounds in the audio (segmentation)
4. Solving a Multi-Instance Multi-Label (MIML) problem

The final model as presented produced very impressive result with the accuracy of 0.8396 and this work can be extended by deep learning techniques.

6.References:

[1] F. Briggs, X. Z. Fern, and J. Irvine,"Multi-label classifier chains for bird sound", arXiv: 1304.5862, (abs/1304.5862), 2013.

[2] F.Briggs, B.Lakshminarayanan, L.Neal, X.Fern, R.Raich, S.J.K.Hadley, A.S.Hadley, and M. G. Betts," Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach", Journal of the Acoustical Society of America, 131:4640–4650, 2012.

[3] F.Briggs, X.Z.Fern,R.Raich,and Q.Lou, "Instance annotation for multi-instance multi-label learning",Transactions on Knowledge Discovery from Data (TKDD), 2012, 2012.

[4] Herrera.F, Charte.F,Rivera A.J, del jesus, "Multi-Label classification", M.J 2016,XVI,194 p. 72 illus,Hardcover.

[5] L. Breiman. Random forests. Machine Learning, 45(1):5–32, 2001.

[6] A.P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7):1145–1159, 1997.

[7] F. Briggs, R. Raich, and X.Z. Fern. Audio classification of bird species: a statistical manifold approach. In 9th IEEE Int. Conf. on Data Mining, pages 51–60, 2009.

[8] Seppo Fagerlund. Automatic recognition of bird species by their sounds. PhD thesis, Helsinki University of Technology, 2004.

[9] T. Scott Brandes. Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conservation Int., 18(S1):163–173, 2008. ISSN 0959-2709.

[10] Z.H. Zhou, M.L. Zhang, S.J. Huang, and Y.F. Li. Multi-instance multi-label learning. Artificial Intelligence, 176(1):2291–2320, 2012.