

Multi-Instance Multi-Label Class Discovery: A Computational Approach for Assessing Bird Biodiversity

Forrest Briggs
Facebook, Inc.
fbriggs@gmail.com

**Xiaoli Z. Fern, Raviv Raich,
Matthew Betts**
Oregon State University
{xfern,raich}@eecs.oregonstate.edu
matt.betts@oregonstate.edu

Abstract

We study the problem of analyzing a large volume of bioacoustic data collected in-situ with the goal of assessing the biodiversity of bird species at the data collection site. We are interested in the class discovery problem for this setting. Specifically, given a large collection of audio recordings containing bird and other sounds, we aim to automatically select a fixed size subset of the recordings for human expert labeling such that the maximum number of species/classes is discovered. We employ a multi-instance multi-label representation to address multiple simultaneously vocalizing birds with sounds that overlap in time, and propose new algorithms for species/class discovery using this representation. In a comparative study, we show that the proposed methods discover more species/classes than current state-of-the-art in a real world dataset of 92,095 ten-second recordings collected in field conditions.

Introduction

Bioacoustic monitoring is a rapidly growing field, where the goal is to learn about organisms such as birds and marine mammals, by applying signal processing and machine learning to audio recordings. In this paper, we consider the problem of class discovery from bird bioacoustics data. Given a large collection of audio recordings of birds (and other sounds in the environment), our goal is to automatically select a subset of recordings to be manually labeled by human expert such that we can find the maximum number of species/classes with a fixed labeling budget.

Acquiring critical knowledge about the response of species to global change necessitates the development efficient and accurate estimates of species abundance and diversity. Birds have been used widely as biological indicators because they respond rapidly to change, are relatively easy to detect, and may reflect changes at lower trophic levels (e.g., insects, plants) (Şekercioğlu, Daily, and Ehrlich 2004). Recently, Wimmer et al. (2013) compared in-field manual point counts, a traditional way for assessing bird biodiversity, to acoustic sampling for discovering bird species. They found that acoustic sampling detected more species for an equivalent amount of human effort. In that study, the recordings are chosen randomly from a particular interval of time

(e.g., 3 hours after dawn), and the audio signal itself was not used to make the selection. In this work, we propose methods for analyzing the audio recordings and for selecting a species-diverse set of recordings for human expert labeling. We apply our proposed methods, and baseline methods, to a real-world dataset of 92,095 ten-second recordings, collected at 13 sites over a period of two months, in a research forest. These recordings pose many challenges for automatic species discovery, including multiple simultaneously vocalizing birds of different species, non-bird sounds such as motor sound, and environmental noises, e.g., wind, rain, streams, and thunder. Our results show that the proposed methods discover more species/classes than previous methods.

Background

Bird bioacoustic data has been considered by the machine learning community primarily for supervised classification tasks. Briggs et al. (2012b) proposed to represent audio recordings of bird sound in the multi-instance multi-label (MIML) framework (Zhou et al. 2012). In this formulation, an audio recording is transformed to a spectrogram, then automatically segmented into a collection of regions believed to be distinct utterances of bird sound. Each segment is then described by a feature vector that characterizes its shape, texture, and time/frequency profiles. A recording is represented as a set of segment feature vectors (instances). Because recordings collected in natural environments often contain sounds from multiple species, each recording is associated with a set of class labels. This multi-instance multi-label representation has previously been successfully used for predicting the bird species present in a recording (Briggs et al. 2012b) and also to predict the single species responsible for a specific utterance (Briggs et al. 2012a) in the recording. In our work, we employ the same representation but the recordings are not labeled. Only after the species discovery algorithm makes its selection, then the selected recordings are given to experts to be labeled.

Bird species discovery from acoustic data has only recently been studied, and prior methods do not make use of the audio data directly, but instead rely on meta data such as the time of day (Wimmer et al. 2013). The class-discovery problem has been studied in the machine learning and data-mining communities, although prior work has fo-

cused on discovering rare classes in single-instance single label data (Pelleg and Moore 2004; He and Carbonell 2007; Vatturi and Wong 2009).

Problem Statement

We consider the species discovery problem in the context of a multi-instance multi-label dataset. Our input is a collection of audio recordings, each represented as a bag of instances. Each recording/bag is associated with a set of species, which are initially unknown, but can be queried.

Formally, the dataset is $(B_1, Y_1), \dots, (B_n, Y_n)$ where B_i is a bag of n_i instances: $B_i = \{\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}\}$, $\mathbf{x}_{ij} \in \mathbb{R}^d$ is a feature vector representing an instance, and $Y_i \subseteq \{1, \dots, s\}$ is a subset of s classes (species or other sounds). We focus on the batch setting where we are given a fixed budget for human effort to label m bags and all m bags must be selected at once. The problem is to select m bags in the absence of any bag label information based only on the information given by the feature value of their instances. A class discovery algorithm picks a set of bag indices R , and is evaluated based on the number of classes discovered $|\cup_{i \in R} Y_i|$.

Species Discovery Methods

In this section, we describe prior methods and introduce some simple yet intuitive baselines, followed by our proposed methods for automatic species discovery.

Prior methods

Wimmer et al. (Wimmer et al. 2013) recently explored temporally stratified methods for species discovery from acoustic monitoring data. In this work, recordings are selected for labeling randomly from within stratified time intervals, without using the recording content to inform the decision. Nonetheless, this work represents the current state-of-the-art in bird species discovery. Different sampling strategies considered in (Wimmer et al. 2013) include random from a full 24-hour period, random from dawn, random from dusk, random from dawn and dusk, and regular intervals. Of these methods, the most effective one is the *dawn* method, which picks random 1-minute recordings from the period from dawn until 3 hours after dawn (often the most active period for birds). In our work, we implement the *dawn* method by selecting m random recordings from 5:00 am to 8:00 am.

Baseline methods

A simple baseline method that has been applied for class discovery (in non multi-instance data) is to cluster all instances, then select one instance from each cluster (e.g., the one closest to the cluster center) (Chen et al. 2013). We take a similar approach for multi-instance data. First, all instances from all bags are clustered with k -means++ (Arthur and Vassilvskii 2007), where the number of clusters k is equal to the number of bags to be selected m . Then, we select the bag containing the instance closest to each cluster center. The clusters are queried in order from most instances to least instances.

Multi-Instance Farthest First (MIFF)

Farthest-first traversal (Gonzalez 1985) is a greedy method that has been successfully applied for class discovery in single-instance single-label data (Chen et al. 2013). It repeatedly selects the instance farthest from the set of currently selected instances to maximize the diversity among the selections. We propose multi-instance farthest first (MIFF) (Algorithm 1), which extends this idea to multi-instance data.

We first describe a basic version of our algorithm, which is a straight forward extension of farthest-first to multi-instance data. It first select a bag randomly from the set of non-empty bags (line 2). Whenever a bag is selected, all of the instances it contains are “covered” (lines 3 and 18). After the first random bag, it repeatedly selects the bag that contains the instance that is farthest from the set of covered instances.

The above basic extension selects the bag based on a single (farthest) instance. It is important to note that for multi-instance multi-label data, querying a bag results in obtaining its full set of labels, rather than a single label. Hence it is advantageous to evaluate a bag by considering more instances and choose a bag with multiple instances that are far from the currently covered instances, and far from each other.

Based on this intuition, we present MIFF with a parameter p that controls the number of instances in each bag that are considered. When $p = 1$, MIFF reduces to the basic version described above. For $p > 1$, a bag is chosen to maximize a sum of p distances that are computed greedily. In each greedy step, we choose the single uncovered instance in the bag that is farthest from the set of covered instances and record this distance. Once an instance is chosen, it is added to the covered set and used to select the next instance in the bag. This process repeats until p instances are chosen and the score of the bag is the sum of the p distances (see lines 10-14).

Cluster Coverage with MIFF (CCMIFF)

To compare the efficiency of different temporally-stratified sampling methods, Wimmer et al. (2013) considered a theoretical estimate of the minimum number of recordings required to detect all species. Assuming all the data are labeled, they applied a greedy algorithm for the classic NP-hard set cover problem to obtain this estimate. We take inspiration from this idea to devise a species discovery method.

Instead of the set cover problem, we consider the related, max cover problem: given a number m , a universe U , and subsets $S_i \subset U, i = 1, \dots, n$, choose m subsets so the size of their union is maximized:

$$\arg \max_{R \subset \{1, \dots, n\}} |\bigcup_{i \in R} S_i| \text{ such that } |R| = m \quad (1)$$

This problem is also NP-hard, but can be approximated by a greedy algorithm that repeatedly selects the S_i that covers the most uncovered elements of U . Feige (1998) proved that this algorithm achieves a $1 - \frac{1}{e}$ approximation ratio, which is the best possible approximation ratio unless $P = NP$.

If the recording are labeled, we can define S_i as the set of species present in recording i . In this case species/classes are

Algorithm 1 Multi-Instance Farthest First (MIFF)

```
1: Input: multi-instance dataset  $\{B_1, \dots, B_n\}$ , number of
   bags to select  $m$ , number of instances per bag to use  $p$ 
2:  $S = \{r\}$  — Initialize  $S$  with a random non-empty bag
    $r$ 
3:  $C = B_r$  —  $C$  stores all covered instances
4: for  $i = 2$  to  $m$  do
5:   — select the  $i$ 'th bag
6:   for  $j = 1$  to  $n$ ,  $j \notin S$ ,  $|B_j| \neq 0$  do
7:     — consider bag  $j$  as a candidate for selection
8:      $v_j = 0$  — a score for bag  $j$ 
9:      $C_j = \{\}$  — the set of instances covered in this bag
10:    for  $l = 1$  to  $p$  do
11:       $\mathbf{x}^* = \arg \max_{\mathbf{x} \in B_j \text{ and } \mathbf{x} \notin C_j} \min D_{nn}(\mathbf{x}, C \cup C_j)$ , where
         $D_{nn}(\mathbf{x}, S) = \min_{\mathbf{x}' \in S} d(\mathbf{x}, \mathbf{x}')$ 
12:       $v_j = v_j + D_{nn}(\mathbf{x}^*, C \cup C_j)$  — update the score
13:       $C_j = C_j \cup \{\mathbf{x}^*\}$  —  $\mathbf{x}^*$  is now covered
14:    end for
15:  end for
16:   $q = \arg \max_j v_j$  — pick the highest scoring bag
17:   $S = S \cup \{q\}$  — add it to the set of selected bags
18:   $C = C \cup B_q$  — update covered instances
19: end for
```

the items in U being covered. However, unlike the retroactive analysis by Wimmer et al., the set of species contained in each recording are unknown to us when we must select the set R . Therefore, we use clusters as a proxy for classes. $U = \{1, \dots, k\}$ is the k clusters obtained by k -means++, and S_i is the set of clusters contained in bag i . We refer to this method as cluster coverage, because clusters are the elements being covered.

Naïvely applying the greedy algorithm, we observe that often many bags are tied for covering the most new clusters. We would prefer to break these ties based on some principle, rather than arbitrarily. Furthermore, before we have selected m bags, we may cover all of the clusters (in which case, all remaining unselected bags are tied with a coverage improvement of 0). We address both of these issues by breaking ties in coverage according to the same criteria used in MIFF. Specifically, instead of evaluating all remaining bags that are non-empty as in MIFF, we only evaluate the bags that are tied for covering the most new clusters and select the bag that has the highest score. This approach is thus referred to as cluster coverage with MIFF tie breaking (CCMIFF).

Experiments

Dataset

In this study, we collected audio data at 13 different sites in the H. J. Andrews Long Term Experimental Research Forest over a two-month period during the 2009 breeding season. The recording devices were programmed to record the first 20 minutes of each hour of the day. A total of 589.75 GB, roughly 2559 hours, of audio recordings was collected, divided into 7,688 WAV files (most are 20 minutes

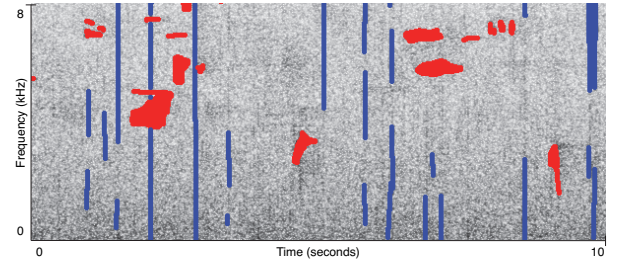
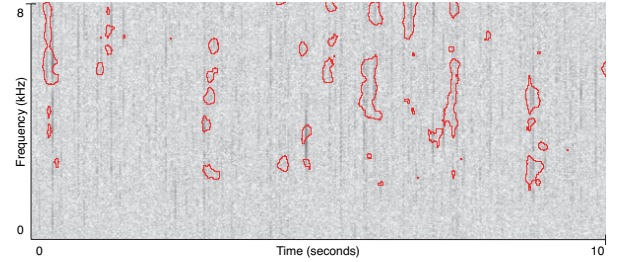
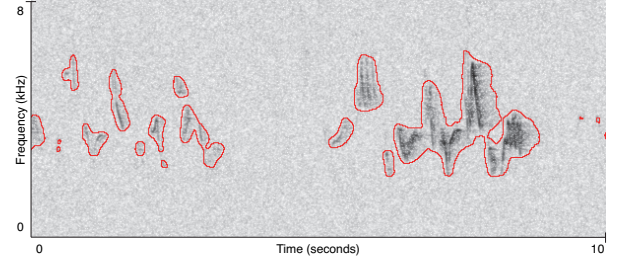


Figure 1: An example of a spectrogram manually annotated with ground-truth segmentation (used to train the automatic segmentation algorithm). Red indicates bird sound, and blue indicates rain or other loud non-bird sound.



(a) Rain Example



(b) Non-Rain Example

Figure 2: Two randomly chosen recordings of rain and non-rain categories, used to train the rain filter. The red outline shows the automatic segmentation.

long). Because it is convenient and efficient to work with smaller intervals of audio, e.g., ten seconds, we divided the full dataset into 920,956 ten-second intervals, then randomly subsampled 10% of this data, to obtain a total of 92,095 tens-second recordings for our experiments. We follow the approach described in (Briggs et al. 2012b) to generate the multi-instance representation of our data. In particular, each recording is first segmented using the algorithm described in (Briggs et al. 2013). Segments/instances are then described by a 38-d feature vector.

Rain Filter

We noted that the spectrogram segmentation algorithm (Briggs et al. 2013) is trained with rain as negative examples, but there are still cases where it fails (particularly when analyzing a large number of recordings). In such cases, the resulting segmentation often consists of many small segments

with a wide variety of shapes (Fig. 2a). These segments tend to confound the species discovery algorithm. To alleviate this problem and avoid selecting rain recordings, we perform recording-level rain filtering using a random forest classifier (Breiman 2001) trained on 1000 ten-second recordings selected randomly from the full dataset and manually labeled as rain/non-rain training examples. Figure 2 shows examples of rain and non-rain recording spectrograms. The input to the classifier is a recording-level feature vector that is computed based on the segments it contains and consists of a histogram of segments, and the mean and standard deviation of the segment-level features. Given a recording, the Random Forest classifier (with 1000 trees) is used to predict the probability of rain. If greater than a threshold T , the recording is removed from consideration. The rain filter can be combined with any of the proposed species discovery methods.

Training Data

Although the proposed methods are primarily unsupervised, some training data were used for segmentation and rain filtering. We annotated 150 randomly chosen ten-second recording spectrograms as examples for segmentation. Figure 1 shows an example of an annotated spectrogram for training the segmentation algorithm. A further 1000 randomly chosen recordings are labeled as rain or non-rain to train the rain filter. The human effort for this labeling task was roughly one hour.

Evaluating Species Discovery Efficiency

To compare the efficiency of our proposed methods, and baseline methods, we conducted the following experiments. From the pool of 92,095 recordings, we apply each of the methods (dawn, cluster centers, MIFF, CCMIFF) to select $m = 100$ recordings to be labeled. We wish to emphasize that the species labels in these recordings are all initially unknown to us. We only discover the labels after the algorithm selects a set to be labeled by an expert. Hence, the expert labeled 1000 ten-second recordings in total for all experiments. Labeling these 1000 recordings required roughly 23 hours of labor. The species discovery algorithms evaluate all 92,095 recordings, however.

We compare the species discovered by cluster centers, MIFF, and CCMIFF with rain filter threshold $T \in \{0.1, 0.01\}$ or no rain filter. For MIFF and CCMIFF, we set the parameter $p = 2$ because we expect on average to have 2 classes per bag. For CCMIFF, we set the number of clusters $k = 1000$, based on the observation that with a smaller number of clusters (e.g., 100), the algorithm covers all clusters very early on, before selecting $m = 100$ bags. Once all clusters are covered, CCMIFF behaves identically to MIFF, so it is only interesting to compare the two with parameters that cause CCMIFF not to cover all clusters right away.

Results

The results of the experiment are viewed in terms of a graph of number of species or classes discovered vs. number of recordings labeled. We construct separate graphs for the count of species (bird species only), and all classes of

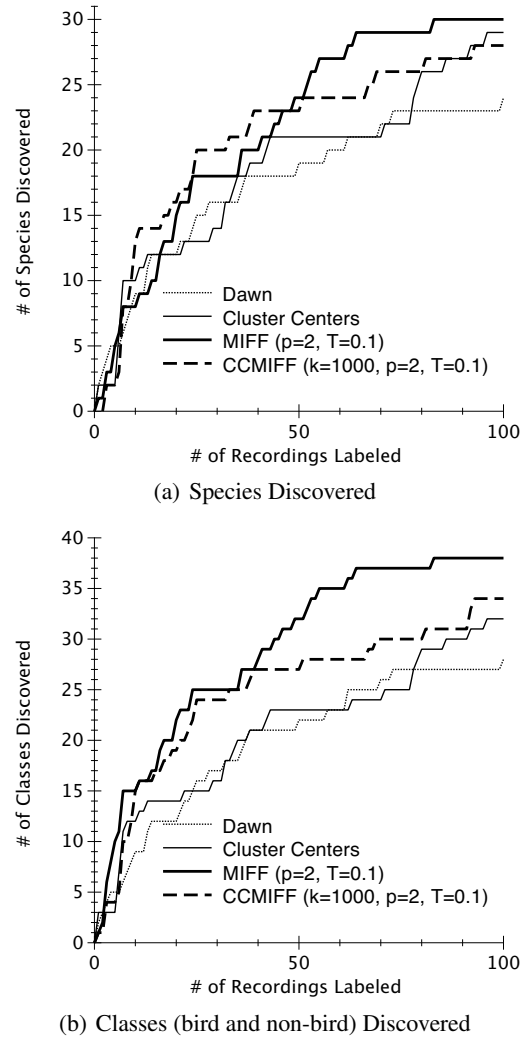


Figure 3: Species/class discovery curves with the best parameters.

sound (e.g., airplanes, thunder, walking, beeps, sticks breaking, etc). Figure 3 shows the number of species and classes discovered by each method with its best parameter setting.

For both species or classes, MIFF with $p = 2$ instances per bag considered, and rain threshold $T = 0.1$ achieved the best result after selecting 100 recordings. However, up to selecting 50 recordings, CCMIFF discovers more species than MIFF, and also achieves the second best results in terms of classes discovered.

Most significantly, all of the methods that use the multi-instance representation of the data (cluster centers, MIFF, and CCMIFF) find more species and classes than the dawn time-based method (Fig. 3). Hence, these results demonstrate progress toward what has been assessed as a very challenging task in (Wimmer et al. 2013).

Figure 4 shows the number of recordings with each label selected by each method. The species/classes are sorted in descending order of their frequency with the dawn method.

Class	Dawn	Cluster Centers	Cluster Centers (T=0.10)	Cluster Centers (T=0.01)	MIFF (k=2)	MIFF (k=2, T=0.10)	MIFF (k=2, T=0.01)	CCMIFF (c=1000, k=2)	CCMIFF (c=1000, k=2, T=0.10)	CCMIFF (c=1000, k=2, T=0.01)
Bird-Species										
SWTH	26	39	43	42	19	30	38	30	38	44
HETH	21	15	15	14	12	25	28	19	28	45
PAWR	18	28	30	29	7	19	19	23	29	14
HEWA	16	20	22	22	8	18	25	17	33	31
HAFL	15	8	9	8	3	3	0	6	4	1
CBCH	11	11	11	11	5	5	10	13	14	7
VATH	10	4	4	6	5	7	17	13	14	20
ORJU	9	3	3	2	5	7	12	4	6	12
WETA	9	6	6	7	1	2	2	3	11	6
WIND	6	2	5	3	19	10	4	5	3	3
OSFL	4	1	2	2	0	1	5	3	4	7
PSFL	4	5	4	4	1	2	7	2	5	5
STJA	4	1	2	1	2	3	2	1	5	1
WAVI	4	5	6	6	3	5	6	3	6	8
MGWA	3	4	4	4	3	3	9	5	8	6
AMRO	2	5	4	3	0	3	0	2	4	4
BTYW	2	0	0	0	0	0	0	1	0	0
RBNU	2	0	0	0	0	3	4	2	1	3
BHGR	1	1	1	1	3	6	12	12	15	15
BRCR	1	1	1	1	1	1	0	1	2	2
CHIP	1	0	0	1	0	0	0	2	1	0
CONI	1	2	2	2	2	3	2	2	2	3
RECR	1	5	5	5	0	2	0	0	0	0
WIWA	1	0	0	0	0	1	0	0	3	0
BTPI	0	0	0	0	0	0	1	0	0	1
CBCB	0	0	0	0	0	0	0	0	0	0
CORA	0	1	1	1	0	1	5	0	0	2
EVGR	0	1	1	1	0	0	0	0	0	0
GCKI	0	4	5	5	0	7	8	3	9	5
HOSA	0	1	1	1	2	1	1	2	2	1
HUVI	0	0	0	0	2	2	0	0	0	0
OSPR	0	0	0	0	0	1	0	0	0	0
PAFL	0	1	0	0	0	0	0	0	0	0
PISI	0	0	0	0	1	0	1	0	1	0
PIWO	0	1	1	1	0	0	0	0	0	0
RBSA	0	0	1	1	0	0	0	0	0	0
SPOW	0	0	0	0	0	0	0	0	0	1
VASW	0	1	0	0	0	0	0	1	0	0
YRWA	0	0	0	0	1	0	0	0	0	0
Other Sounds										
RAIN	3	13	15	12	24	6	4	22	6	3
ARPL	1	13	15	15	11	19	16	5	12	11
MOTOR	1	0	0	0	5	9	0	6	4	4
DOSQ	1	1	0	0	3	6	3	4	3	2
THUNDER	0	0	0	0	0	1	0	0	0	0
INCT	0	5	5	5	8	9	8	7	9	9
WALK	0	0	0	0	2	0	0	0	0	0
WATER	0	0	0	0	0	0	0	0	1	0
TAPP	0	1	0	0	0	0	0	0	0	0
SMGLITCH	0	0	0	0	12	4	0	3	0	0
OTHR	0	0	0	0	0	1	0	0	1	0
MICBUMP	0	0	0	0	4	1	0	2	1	0
HAMMER	0	0	0	0	1	3	0	0	0	0
BEEP	0	0	0	0	2	2	0	1	1	0
BREAK	0	0	0	0	1	0	0	0	0	0

Figure 4: The number of recordings of each class selected by each method.

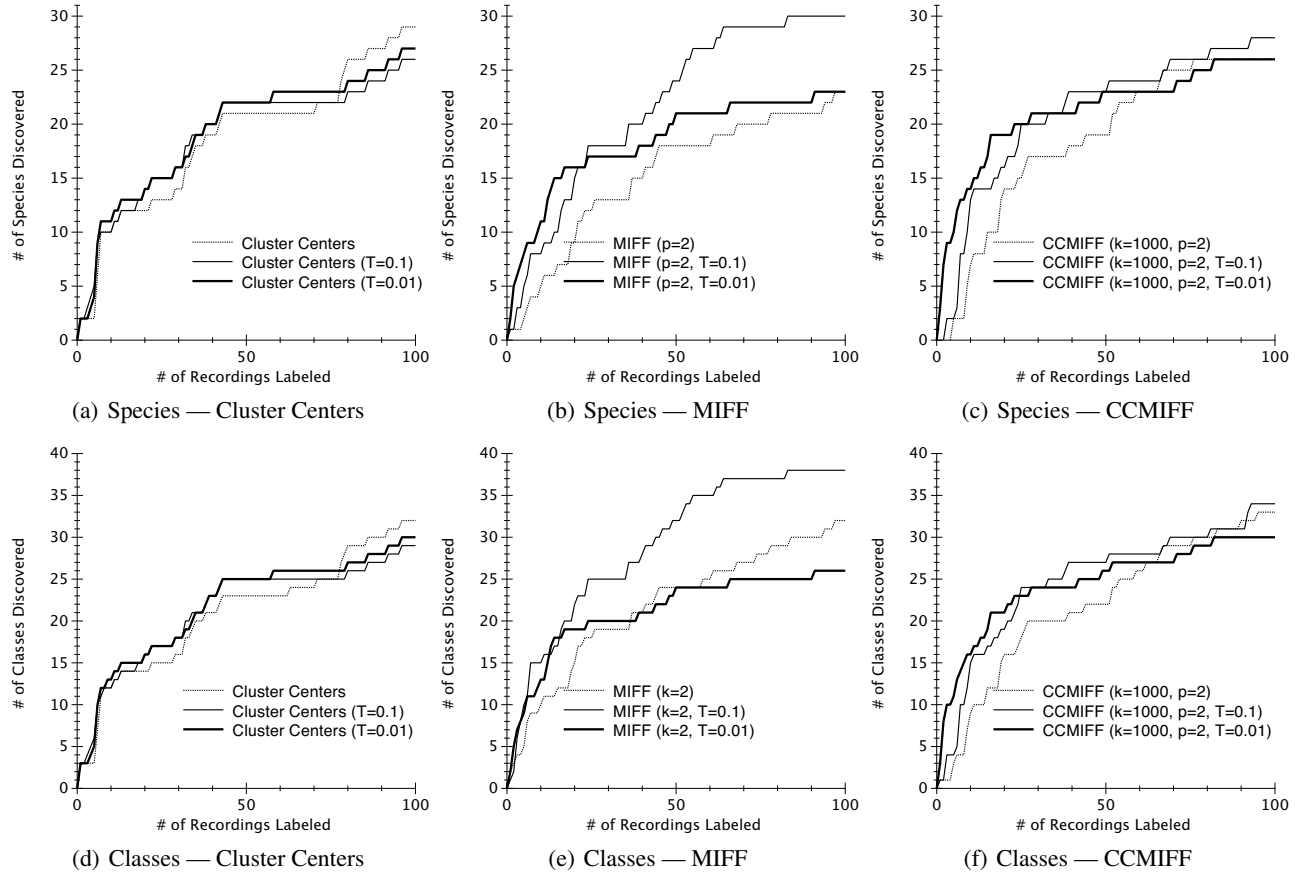


Figure 5: Sensitivity of species and class discovery curves to varying rain probability threshold T .

Table 1: Class codes and descriptions for non-bird sounds.

Code	Description
RAIN	rain drops
ARPL	airplane motor
MOTOR	other motor vehicle
DOSQ	Douglas Squirrel
THUNDER	thunder
INCT	insect buzzing
WALK	a person walking near the Songmeter
WATER	water flowing
TAPP	woodpecker tapping sound
SMGLITCH	Songmeter glitch
OTHR	unknown
MICBUMP	microphone being bumped
HAMMER	hammer strike
BEEP	the Songmeter or a watch beeping
BREAK	a stick snapping

The first group of labels are bird species only, identified by a standard 4-letter code (Union 1910). The second group of labels is for non-bird sounds (Table 1). This chart shows that for cluster centers, MIFF, and CCMIFF without the rain filter, many of the selected recordings include

the RAIN label. We hypothesized that the number of other species/classes discovered could be improved by selecting fewer rain recordings. Table 1 also shows that as the rain threshold parameter T decreases, so does the number of rain recordings selected by all methods.

Figure 5 shows the sensitivity of cluster centers, MIFF, and CCMIFF to the rain threshold parameter T . The cluster centers method finds the most classes after labeling 100 recordings with no rain filter, although for lower numbers of labeled recordings, the most restrictive rain filter parameter $T = 0.01$ gives better results. MIFF and CCMIFF achieve best results in terms of both species and classes with $T = 0.1$. Hence, we see that the rain filter generally provides some benefit for these two algorithms (by preventing too many queries from being wasted on rain).

Discussion

In this paper, we addressed the problem of selecting a subset of audio recordings from a large dataset to be labeled by an expert, to maximize the number of species discovered for a fixed amount of human effort. Previous state-of-the-art methods in this application used only the time meta data to select recordings. In contrast, our proposed methods analyze the audio content of the recordings to improve

the selection, and successfully handle many of the complexities of real world data such as multiple simultaneous birds, rain, and other non-bird sounds. Experiments suggest that our proposed methods discover species more efficiently than time stratified acoustic sampling (which has previously been shown to be more efficient than traditional point counts).

Biodiversity is a critical indicator of ecosystem health and an important factor to consider in conservation management. Traditional biodiversity surveys requires experienced birders to conduct in-field point counts, which are time consuming, challenging or impractical in remote areas, and often miss rare species. The key advantages of our method are that it makes more efficient use of human effort to measure biodiversity, and can provide better temporal coverage, which improves detection of rare species.

In future work, we will use the labeled recordings obtained in this study as a training set for a supervised classifier. This classifier will then be applied to the full 2009 dataset to predict the species for all sounds in the dataset. With these predictions, we expect to be able to identify ecologically interesting patterns in activity and phenology.

Acknowledgements

This work was supported in part by the National Science Foundation grants IIS-1055113, CCF-1254218, and DBI-1356792.

References

- Arthur, D., and Vassilvitskii, S. 2007. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 1027–1035. Society for Industrial and Applied Mathematics.
- Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.
- Briggs, F.; Fern, X.; Raich, R.; and Lou, Q. 2012a. Instance annotation for multi-instance multi-label learning. *Accepted pending revision, Transactions on Knowledge Discovery from Data (TKDD)*, 2012.
- Briggs, F.; Lakshminarayanan, B.; Neal, L.; Fern, X.; Raich, R.; Hadley, S.; Hadley, A.; and Betts, M. 2012b. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America* 131:4640.
- Briggs, F.; Huang, Y.; Raich, R.; Eftaxias, K.; Lei, Z.; Cukierski, W.; Hadley, S. F.; Hadley, A.; Betts, M.; Fern, X. Z.; et al. 2013. The 9th annual mlsp competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, 1–8. IEEE.
- Chen, Y.; Groce, A.; Zhang, C.; Wong, W.-K.; Fern, X.; Eide, E.; and Regehr, J. 2013. Taming compiler fuzzers. In *PLDI*, 197–208.
- Feige, U. 1998. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM (JACM)* 45(4):634–652.
- Gonzalez, T. F. 1985. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science* 38:293–306.
- He, J., and Carbonell, J. G. 2007. Nearest-neighbor-based active learning for rare category detection. In *Advances in neural information processing systems*, 633–640.
- Pelleg, D., and Moore, A. W. 2004. Active learning for anomaly and rare-category detection. In *Advances in Neural Information Processing Systems*, 1073–1080.
- Şekercioğlu, Ç. H.; Daily, G. C.; and Ehrlich, P. R. 2004. Ecosystem consequences of bird declines. *Proceedings of the National Academy of Sciences* 101(52):18042–18047.
- Union, A. O. 1910. *Check-list of North American birds*. American Ornithologists' Union.
- Vatturi, P., and Wong, W.-K. 2009. Category detection using hierarchical mean shift. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 847–856. ACM.
- Wimmer, J.; Towsey, M.; Roe, P.; and Williamson, I. 2013. Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications*.
- Zhou, Z.-H.; Zhang, M.-L.; Huang, S.-J.; and Li, Y.-F. 2012. Multi-instance multi-label learning. *Artificial Intelligence* 176(1):2291–2320.