Multi-Label Classifier Chains for Bird Sound

Forrest Briggs BRIGGSF@EECS.OREGONSTATE.EDU
Xiaoli Z. Fern XFERN@EECS.OREGONSTATE.EDU
Jed Irvine IRVINE@EECS.OREGONSTATE.EDU
Oregon State University, Corvallis, OR, 97333, USA

Abstract

Bird sound data collected with unattended microphones for automatic surveys, or mobile devices for citizen science, typically contain multiple simultaneously vocalizing birds of different species. However, few works have considered the multi-label structure in birdsong. We propose to use an ensemble of classifier chains combined with a histogram-of-segments representation for multi-label classification of birdsong. The proposed method is compared with binary relevance and three multi-instance multi-label learning (MIML) algorithms from prior work (which focus more on structure in the sound, and less on structure in the label sets). Experiments are conducted on two real-world birdsong datasets, and show that the proposed method usually outperforms binary relevance (using the same features and base-classifier), and is better in some cases and worse in others compared to the MIML algorithms.

1. Introduction

The most familiar formulation of supervised classification associates single feature-vectors with single labels, hence it is called single-instance single-label (SISL). For example, SVM and logistic regression are SISL classifiers. One common setup involving SISL classifiers is to use a segmentation algorithm to extract "syllables" or calls of bird sound from a recording, each of which is described by a feature vector. A SISL classifier is trained on a collection of syllables paired with species labels, then predicts the species for a new syllable (Fagerlund, 2007; Damoulas et al., 2010).

Many of the audio recordings used in SISL experiments are collected with a directional microphone aimed by a person at the bird of interest. This method produces recordings where the targeted bird is louder than other sound sources in the environment. Audio data collected by unattended microphones for the purpose of acoustic monitoring, and audio collected with mobile devices for citizen science are less ideal; it is common to have multiple simultaneously vocalizing bird species, in addition to other sources of noise such as non-bird species, wind, rain, streams, and motor vehicles. Few works have addressed these complexities in real-world data (Brandes, 2008; Briggs et al., 2012c).

There are two kinds of structure in bird sound data that can be exploited through alternative frameworks for supervised classification. First, bird sound is naturally decomposed into a collection of parts, e.g., syllables, which motivates a multi-instance learning (MIL) approach (Dietterich et al., 1997). Second, multi-label classification (MLC) (Tsoumakas & Katakis, 2007) is a natural fit for bird sound because an audio recording can be associated with a set of species (and other sounds) that are present. Multi-instance multi-label learning (MIML) combines both ideas. MIML has previously been used for classification of bird sound recordings containing multiple simultaneously vocalizing species (Briggs et al., 2012c). However, prior work on MIML for bird sound has focussed more on the multi-instance structure of the sound, and less on structure in the species/label sets.

The MLC framework has not been directly applied to bird sound (although some MIML algorithms which have been applied to bird sound can be considered a reduction to MLC, e.g., MIML-kNN (Zhang, 2010) and MIML-RBF (Zhang & Wang, 2009)). Ensemble of classifier chains (ECC) (Read et al., 2011) is an algorithm for MLC which has recently been applied to species distribution modeling, where the goal is to predict the set of bird species present at a site from a feature vector describing physical and biological properties of the site. Yu et al. (Yu et al., 2011) suggested

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28. Copyright 2013 by the author(s).

that ECC achieves better performance in this domain than binary relevance because it can exploit correlations in the label sets. Considering this observation, we hypothesize that ECC can exploit the same structure while predicting sets of bird species from an acoustic feature vector instead of environmental covariates.

We formulate the classification problem similarly to (Briggs et al., 2012c). The training data consists of audio recordings paired with a set of species that are present. The goal is to predict the set of species in a new recording which is not part of the training data.

To apply MLC, it is necessary to represent each audio recording with a fixed-length feature vector. We apply a 2D time-frequency supervised segmentation algorithm similar to (Neal et al., 2011; Briggs et al., 2012c), then compute the same features as in (Briggs et al., 2012c) to describe each segment. Then we use a clustered codebook to obtain a histogram-of-segments for each recording. (Somervuo & Harma, 2004) used histograms to represent variable-length sequences of syllables. (Briggs et al., 2009) used histograms of frame-level features (spectrum and MFCC) to represent an audio recording with a single species of bird.

We compare ECC, binary relevance (BR), and results from prior work on two real-world datasets of birdsong with multiple simultaneously vocalizing species.

The first dataset was collected with unattended omnidirectional microphones in the H. J. A. (HJA) Experimental Research Forest, and has previously been used in several classification experiments (Briggs et al., 2012c;a;b; Liu & Dietterich, 2012)

The second dataset is new, and consists of recordings of birds made with an iPhone in a residential neighborhood (collected and labeled by the authors). The new iPhone birdsong dataset presents the same multi-species issues as the HJA Birdsong dataset, but is arguably more challenging because there are more/louder sources of background noise and non-bird classes (especially motor vehicles and insects).

Results are analyzed in terms of standard multi-label error measures: Hamming loss, set 0/1 loss, rank loss, 1-error, and coverage. ECC achieves better results than BR in the majority of comparisons, and ECC with no parameter tuning is better than one and worse than two of the MIML algorithms (which have an unfair advantage of using post-hoc parameter tuning).

2. Problem Statement

In MLC, the training dataset is $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ where $\mathbf{x}_i \in \mathbb{R}^d$ is a feature vector, and $Y_i \subseteq \mathcal{Y} = \{1, \ldots, c\}$ is a subset of c possible class labels. The goal is to learn a classifier $f(\mathbf{x}) : \mathbb{R}^d \to 2^{\mathcal{Y}}$ which predicts a label set from a given feature vector. It is common to implement and evaluate multi-label classifiers based on a score function for each class $f_j(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$, which represents the predicted confidence that label j is in the set. The set predictor f is defined in terms of the score functions $f_1, \ldots, f_c$. The MLC framework maps to acoustic species classification as follows: each audio recording is associated with a feature vector, and the set of species audible in the recording is the label set.

MIML is a related framework where the training data consists of bags-of-instances paired with label sets,

$$(B_1, Y_1), \ldots, (B_n, Y_n) \text{ where } B_i = \{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\} \quad (1)$$

We will use MIML as an intermediate representation of audio recordings of bird sound, and solve the problem by a reduction from MIML to MLC.

3. Background

Binary relevance is one of the simplest algorithms for MLC. It is a reduction to SISL where binary prediction of each label is treated as a completely separate/independent problem. To refer to a bit in the binary representation of a label set, let $Y_i^j = I[j \in Y_i]$. BR creates c SISL datasets $D_1, \ldots, D_c$, where $D_j = \{(\mathbf{x}_i, Y_i^j)\}_{i=1}^n$, and trains a binary SISL classifier $f_j : \mathbb{R}^d \to \mathbb{R}$ on each D_j.

Classifier chains are also a reduction to SISL, but the problems for each class are not totally separate. CC predicts bits of the label set one at a time in a particular order, and uses all of the previously predicted bits as features for the next bit. CC creates c SISL datasets $D_1, \ldots, D_c$, where

$$D_j = \{(\mathbf{x}_i \oplus Y_i^{1:j-1}, Y_i^j\}_{i=1}^n \quad (2)$$

The notation $Y_i^{1:j-1}$ denotes the first $j-1$ bits of the binary representation of Y_i, and $\oplus$ is vector concatenation. CC trains a binary SISL classifier $f_j : \mathbb{R}^{d+j-1} \to \mathbb{R}$ on each dataset D_j. Algorithm 1 is pseudocode for classification of a feature vector $\mathbf{x}$ with CC. Assuming the SISL classifier f_j outputs a score or probability, a threshold t is used to make a 0/1 prediction.

ECC creates an ensemble of L classifier chains, where each chain $l = 1, \ldots, L$ views the classes in a different random permutation $\pi_l : \{1, \ldots, c\} \to \{1, \ldots, c\}$. Each chain in the ensemble votes on each potential class in the label set. For each chain l and class j, ECC trains a SISL classifier f_{lj} on the dataset

$$D_{lj} = \{(\mathbf{x}_i \oplus Y_i^{\pi_l(1)} \oplus \ldots \oplus Y_i^{\pi_l(j-1)}, Y_i^{\pi_l(j)}\}_{i=1}^n \quad (3)$$

Algorithm 1 Classifier Chains – classify $\mathbf{x}$

```
$Y = []$
for $j = 1$ to $c$ do
  $Y = Y \oplus I[f_j(\mathbf{x}_i \oplus Y) > t]$
end for
return $Y$
```

Algorithm 2 ECC-RF – class scores for $\mathbf{x}$

```
$score[1,\ldots,c] = 0$
for $l = 1$ to $L$ do
  $\mathbf{x}' = \mathbf{x}$
  for $j = 1$ to $c$ do
    $p_{lj} = f_{lj}(\mathbf{x}')$
    $score[\pi_l(j)] = score[\pi_l(j)] + p_{lj}$
    if $j \neq 1$ then
      $\mathbf{x}' = \mathbf{x}' \oplus p_{lj}$
    end if
  end for
end for
return $scores/L$
```

4. Proposed Methods

4.1. Classifier Chains with Random Forest

We implement ECC with a Random Forest (RF) as the base-SISL classifier, hence we call the proposed classifier ECC-RF. Because RF outputs a probability, the ensemble can be viewed as an instance of the Ensemble of Probabilistic Classifier Chains (EPCC) algorithm (Dembczynski et al., 2010). Therefore it is reasonable to aggregate probabilities from each SISL classifier rather than 0/1 votes. The aggregated probabilities are used as the score-functions for each class. Algorithm 2 gives pseudocode we use to generate a class-score vector with ECC-RF, given input $\mathbf{x}$.

4.2. Out-Of-Bag Calibrated Thresholds

Sometimes class scores are sufficient, for example to rank species from most likely to least likely to be present. However, it is often desirable to obtain a specific predicted label set. A label set can be obtained by comparing each score to a threshold. The simplest method is to use a single threshold for all classes (Tsoumakas & Katakis, 2007). We instead select a separate threshold for each class, which is calibrated using out-of-bag (OOB) estimation (Breiman, 2001) (for both BR and ECC-RF). Consider one of the binary RF's in BR or ECC-RF, f_j or f_{lj}. Let its OOB estimate on instance $\mathbf{x}_i$ in the training dataset be $\hat{f}_j(\mathbf{x}_i, i)$ (for BR) or $\hat{f}_{lj}(\mathbf{x}_i, i)$ (for ECC). For each class j, we select a threshold t_j to minimize the 0/1 error on that class, comparing ground-truth labels for class j with OOB estimates. The threshold used in BR for class j is

$$t_j = \operatorname*{arg\,min}_{t \in \{.001,\ldots,.999\}} \sum_{i=1}^{n} I[I[\hat{f}_j(\mathbf{x}_i, i) > t] = Y_i^j] \qquad (4)$$

The same algorithm is applied to ECC-BR by defining $\hat{f}_j = L^{-1} \sum_{l=1}^{L} \hat{f}_{jl}$.

5. Experiments

5.1. Datasets

Two real-world birdsong datasets are used in our experiments.

HJA Birdsong The HJA Birdsong dataset consists of 548 ten-second audio recordings collected in the H. J. A. Experimental Research Forest, using Songmeter SM1 recording devices. There are 13 species in this dataset, with between 1 and 5 species per recording (2.144 average). The most common sources of noise in this dataset include streams and wind. Further details of this dataset are available in (Briggs et al., 2012c). (Briggs et al., 2012c) used 5-fold cross-validation for this dataset. We use the same 5-fold partitions, so the results are comparable.

iPhone Birdsong We collected 150 five-second audio recordings of bird sound with an iPhone 4G in a residential neighborhood. 54 of the recordings were collected during the dawn chorus on a single day, and the rest were collected at different times of day over several months in 2012–13.

We filtered the original 150 recordings down to 91 which are more suited for a cross-validated species classification experiment. There were 32 recordings with bird species we were unable to identify, and many more with non-bird sounds. We removed all recordings containing unknown bird species, amphibians, human voice, dogs barking, and the iPhone vibrating due to receiving a message. Finally, we remove all recordings containing a species which appears only once in the dataset (cross-validation is not reasonable in this case). The filtered subset of 91 recordings contains 14 species. Many of these recordings still contain motor vehicle noise, loud insects, and "click noises" which appear as vertical lines in the spectrogram. Table 1 lists each species, and the number of recordings it appears in. Note that the dataset is highly unbalanced. Because this is smaller dataset, we use 10-fold cross-validation instead of 5-fold.

Table 1. The number of recordings containing each species in the iPhone Birdsong dataset.

Species	Recordings
American Goldfinch	2
American Robin	23
Black Capped Chickadee	36
Black-headed Grosbeak	2
Chestnut Backed Chickadee	3
Golden Crowned Kinglet	6
Great Horned Owl	2
Killdeer	7
Marsh Wren	3
Northern Flicker	4
Red Breasted Nuthatch	19
Red-Winged Blackbird	23
Spotted Towhee	13
Stellar's Jay	4

5.2. Histogram-of-Segments Representation

In order to apply MLC, we represent each audio file with a fixed-length feature vector. Prior work (Briggs et al., 2012c) has shown that 2D time-frequency segmentation of a spectrogram is useful for separating bird sounds which may overlap in time. For the new iPhone Birdsong dataset, we follow a similar process to (Briggs et al., 2012c) for supervised 2D segmentation of spectrograms.[1]

Each segment is isolated, and described by the same 38 acoustic features as in (Briggs et al., 2012c). At this point, the audio dataset is represented as a MIML dataset (each recording is a bag of segments paired with a set of species). We reduce this MIML dataset to an MLC dataset by summarizing all of the segments in a recording with a histogram. Hence, the feature vector used for MLC has dimension k, where k is the number of clusters. For the HJA Birdsong dataset, we use the original segmentation and segment features from (Briggs et al., 2012c), rather than our slightly modified segmentation.

Segment features are clustered using k-means++ (Arthur & Vassilvitskii, 2007) to form a codebook. For each recording, each of its segments is mapped to a cluster center, and the normalized count of segments for each cluster is used as the histogram-of-segments feature. Figure 1 shows some example clusters from the codebook for the iPhone dataset.

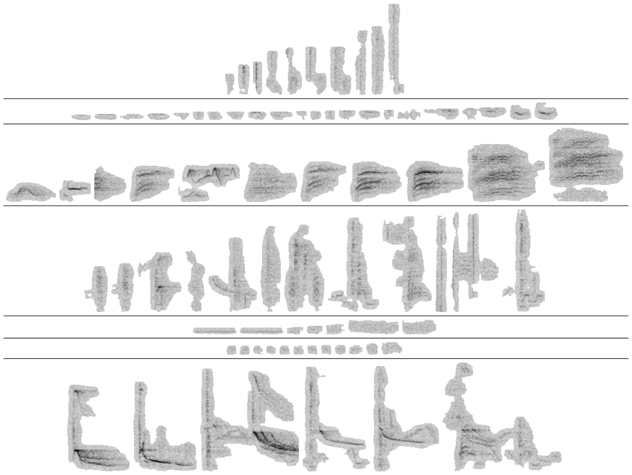

Figure 1. Example clusters of segments in the codebook used in the construction of histogram-of-segment features for the iPhone dataset (modified to enhance contrast).

5.3. Comparison to MIML

Using results from (Briggs et al., 2012c) on the HJA dataset, we compare our proposed ECC-RF algorithm to three MIML algorithms: MIMLSVM, MIML-kNN and MIMLRBF. Each of these algorithms are reductions from MIML to MLC; they construct a single fixed-length feature vector from a bag of instances (i.e., a recording containing a varying number of segments), then apply binary relevance. For BR, MIMLSVM uses SVM as the base-SISL classifier, while MIML-kNN and MIMLRBF use linear models trained by unregularized min-squared-error. These MIML classifiers focus mainly on construction of a good "summary" feature vector, while using only the simplest MLC classifier. In contrast, our proposed method uses a simpler feature vector construction, and a more complicated model of structure in the label sets.

5.4. Parameters

For constructing histogram of segment features, the parameter to k-means++ is $k = 50$.

The only parameters for ECC-RF are L, the number of chains, and T, the number of trees in each RF. It is expected that as these parameters are increased, the accuracy of the classifier converges to some asymptotic value. Hence selection of these parameters is mainly a matter of how much computation time is available. We conservatively chose $L = 25, T = 25$, and did no further optimization of these parameters.[2]

[1] There are some minor differences in segmentation in the iPhone dataset vs. the HJA dataset. For the iPhone dataset, the RF used for segmentation was trained on features consisting of pixels in an 17x17 window, the y-coordinate of the window center, and the average intensity in the window. This RF used 100 trees with a maximum depth of 10. We annotated 20 out of 91 of the spectrograms in the dataset with examples of correct segmentation.

[2] Running 10 repetitions of 5- or 10-fold CV on both datasets with BR and ECC-RF takes 424 seconds on a Mac Pro with 2x2.4 GHz Quad-Core Intel Xeon Processors. The

For BR, the only parameter is T, the number of trees in each RF. We set $T = 25^2$ for BR to ensure that the total number of trees which cast a vote in every prediction is the same between BR and ECC-RF. All decision trees used in both BR and ECC use a maximum tree-depth of 15, and store histograms of class labels in decision tree leaves instead of the majority label.

The three MIML algorithms that we compare to in this experiment have parameters which must be tuned (e.g., by grid search). These tuning parameters are unlike the parameters of ECC-RF. Although such parameters can be optimized by cross-validation (with respect to a particular multi-label performance measure), doing so adds an order of magnitude runtime to the classification experiment, so (Briggs et al., 2012c) used "post-hoc" parameter selection. In post-hoc selection, the experiment is run multiple times for all combinations of parameter values in a grid, and the best result from any parameter is reported. Therefore the MIML algorithms have an advantage in these experiments.

5.5. Results

Table 2 lists results. Because RF and ECC are randomized, we run 10 trials, and report results averaged over all trials and folds of cross-validation.

Following recommendations in (Demšar, 2006), we summarize results for multiple classifiers on multiple datasets by win-loss counts (and do not discard any result as "insignificant"). However, unlike the scenario considered by (Demšar, 2006), we compare MLC classifiers rather than SISL classifiers, so there are multiple performance measures. Because there are only a few datasets and more performance measures, we aggregate win/loss counts over all measures.

Comparing BR and ECC-RF on two datasets with five different performance measures gives 10 comparisons between the two algorithms. Over both datasets, the win-loss count for ECC-RF vs. BR is 7-3. On the iPhone dataset, the result is less decisive; the count for ECC-RF vs. BR is 3-2. On the HJA Birdsong dataset, the count for ECC-RF vs. BR is 4-1. Overall these results suggest there is an advantage to using ECC-RF over BR for multi-label classification of bird species sets, given the histogram-of-segments representation.

Next we consider the win-loss counts on the HJA Birdsong dataset for ECC-RF vs. MIMLSVM, MIMLRBF, and MIML-kNN. The counts are 5-0, 1-4, and 0-5, respectively, i.e. MIMLSVM is worse than ECC-RF in all comparisons, but MIMLRBF and MIML-kNN are better than ECC-RF. However, this is not an entirely fair comparison due to post-hoc parameter selection in the MIML experiments.

6. Discussion

We suggest that the performance advantage of MIMLRBF and MIML-kNN over ECC-RF may be attributed to better representation of the multi-instance structure in the data (compared to our histogram-of-segments representation). Based on comparisons between ECC and BR, better modeling of structure in the label set is beneficial when compared with the same features and base-SISL classifier.

7. Related Work

We focussed on learning to predict species label sets. Another interesting problem is to train on recordings with multiple labels, but classify segments with a single label. Such an approach reduces the labeling effort required to train SISL segment/syllable classifiers such as (Fagerlund, 2007; Damoulas et al., 2010). This problem is naturally formulated in the framework of MIML instance annotation (Briggs et al., 2012a;b). A related formulation is to associate each segment with a set of candidate labels, only one of which is correct. This formulation is called ambiguous label classification (Cour et al., 2011), or superset label learning (Liu & Dietterich, 2012).

References

Arthur, David and Vassilvitskii, Sergei. k-means++: The advantages of careful seeding. In *Proc. 18th ACM-SIAM symposium on Discrete algorithms*, pp. 1027–1035. Society for Industrial and Applied Mathematics, 2007.

Brandes, T Scott. Feature vector selection and use with hidden markov models to identify frequency-modulated bioacoustic signals amidst noise. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(6):1173–1180, 2008.

Breiman, L. Random forests. *Machine learning*, 45(1): 5–32, 2001.

Briggs, F., Fern, X.Z., and Raich, R. Rank-loss support instance machines for miml instance annotation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 534–542. ACM, 2012a.

RF tree induction is parallel and the rest is sequential. The implementation is in C++ compiled with GCC 4.2.

Table 2. Multi-label classification experiment results, averaged over 10 repetitions.
† – Results from (Briggs et al., 2012c) using post-hoc parameter selection.

Dataset	Classifier	Hamming loss ↓	Set 0/1 Loss ↓	Rank loss ↓	1-error ↓	Coverage ↓
iPhone Birdsong	BR-RF	0.1148	0.811	0.1948	0.5154	3.5495
iPhone Birdsong	ECC-RF	0.1168	0.8121	0.1927	0.5132	3.5319
HJA Birdsong	BR -RF	0.0489	0.4476	0.0258	0.044	1.6805
HJA Birdsong	ECC-RF	0.0485	0.4369	0.0246	0.0482	1.6555
HJA Birdsong	MIMLSVM †	0.054	N/A	0.033	0.067	1.844
HJA Birdsong	MIMLRBF †	0.049	N/A	0.022	0.034	1.632
HJA Birdsong	MIML-kNN †	0.039	N/A	0.019	0.036	1.589

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

Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J.K., Hadley, A.S., and Betts, M.G. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131:4640, 2012c.

Briggs, Forrest, Raich, Raviv, and Fern, Xiaoli Z. Audio classification of bird species: a statistical manifold approach. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pp. 51–60. IEEE, 2009.

Cour, T., Sapp, B., and Taskar, B. Learning from partial labels. *Journal of Machine Learning Research*, 12:1225–1261, 2011.

Damoulas, Theodoros, Henry, Samuel, Farnsworth, Andrew, Lanzone, Michael, and Gomes, Carla. Bayesian classification of flight calls with a novel dynamic time warping kernel. In *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, pp. 424–429. IEEE, 2010.

Dembczynski, Krzysztof, Cheng, Weiwei, and Hüllermeier, Eyke. Bayes optimal multilabel classification via probabilistic classifier chains. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 279–286, 2010.

Demšar, J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30, 2006.

Dietterich, T.G., Lathrop, R.H., and Lozano-Pérez, T. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2): 31–71, 1997.

Fagerlund, Seppo. Bird species recognition using support vector machines. *EURASIP Journal on Advances in Signal Processing*, 2007, 2007.

Liu, Liping and Dietterich, Thomas. A conditional multinomial mixture model for superset label learning. In *Advances in Neural Information Processing Systems 25*, pp. 557–565, 2012.

Neal, L., Briggs, F., Raich, R., and Fern, X. Time-frequency segmentation of bird song in noisy acoustic environments. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 2011.

Read, Jesse, Pfahringer, Bernhard, Holmes, Geoff, and Frank, Eibe. Classifier chains for multi-label classification. *Machine learning*, 85(3):333–359, 2011.

Somervuo, Panu and Harma, Aki. Bird song recognition based on syllable pair histograms. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 5, pp. V–825. IEEE, 2004.

Tsoumakas, G. and Katakis, I. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3):1–13, 2007.

Yu, Jun, Wong, Weng-Keen, Dietterich, Tom, Jones, Julia, Betts, Matthew, Frey, Sarah, Shirley, Susan, Miller, Jeffery, and White, Matt. Multi-label classification for species distribution modeling. In *Proc. ICML 2011 Workshop on Machine Learning for Global Challenges*, 2011.

Zhang, Min-Ling. A k-nearest neighbor based multi-instance multi-label learning algorithm. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, volume 2, pp. 207–212. IEEE, 2010.

Zhang, M.L. and Wang, Z.J. MIMLRBF: RBF neural networks for multi-instance multi-label learning. *Neurocomputing*, 72(16-18):3951–3956, 2009.