Investigation of domain adaptation for acoustic frog species classification

Jie Xie

University of Waterloo, Waterloo, Canada
j65xie@uwaterloo.ca

Abstract

Acoustic frog species classification has received much attention for its importance in assessing biodiversity. However, most previous frog call classification models are trained and tested using the data collected from the same area, which greatly limits the model's generalization. In practice, frogs often have regional accents. When training and testing data are collected from different areas, there is an adverse impact on frog call classification performance. To tackle this problem, this paper investigates domain adaptation for classifying frog calls collected from different areas. To evaluate the performance of our proposed methods, two frog call datasets, which are collected from subtropical eastern Australia and tropical north-eastern Australia, are used. Experimental results demonstrate that domain adaptation can significantly improve the weighted F1-score from 72.8% to 85.5%.

1 Introduction

Acoustic classification of frogs has received much attention for its promising applications in assessing the biodiversity. Traditional methods for classifying frog calls are often based on a frog listener, who stands around a pound and listens to frog calls for certain minutes. Then, the listener subjectively assesses how many frog species are there in that specific area. Recently, advances in recording and storage technology provide a novel way to classify frog species. Compared to traditional field survey methods, the use of acoustic sensors greatly extend the spatiotemporal monitoring scale [13]. Therefore, large volumes of acoustic data have been collected from various locations. When training and testing data are collected from different areas, there will be an adverse impact on frog call classification performance.

Many authors have proposed various methods for acoustic frog species classification [8, 15, 14], but more work is needed to address the problem of classifying frog calls collected from different areas. Most previous classification schemes have an assumption that the probability distributions of the training and testing datasets are similar. However, this assumption is reasonable only when training and testing datasets are collected from the same area. Unlike prior work in automatic frog call classification, in this paper, we consider the following classification problem: frog recordings, which are used as the training and testing datasets, are collected from different areas.

We investigate domain adaptation for acoustic frog species classification. To the best of knowledge, this paper is the first attempt to investigate domain adaptation in the bioacoustics classification problem. Two frog call datasets, which are collected from Subtropical Eastern Australia and Tropical North-eastern Australia, are used. To be specific, frog recordings are first segmented into individual syllables, where a novel acoustic feature set is constructed. Then, domain adaptation is applied to the build feature set to reduce the discrepancy between the source and target domain. Finally, support vector machines (SVMs) are used for the classification with the adapted feature set.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26485v1 | CC BY 4.0 Open Access | rec: 3 Feb 2018, publ: 3 Feb 2018

The remaining part of this paper is organized as follows: Section 2 describes the proposed methods in detail and analyzes their characteristics. Experiments results are shown in Section 3. Conclusions are drawn in Section 4 along with the future research.

2 Methods and Experiments

The flow diagram of the proposed approach is shown in Fig. 1, which can be divided into five modules: pre-processing, syllable segmentation, feature extraction, domain adaptation, and classification.

Figure 1: The flow diagram of our proposed approach.

Table 1: Two frog call dataset sets. Here, SET and TNE denote Subtropical Eastern Australia and Tropical North-eastern Australia.

Species Name	Acronym	SET Syllables No.	TNE Syllables No.	Species Name	Acronym	SET Syllables No.	TNE Syllables No.
Adelotus brevis	ASS	7	4	*Litoria caerulea*	LIE	84	90
Crinia deserticola	CAA	83	85	*Litoria chloris*	LCS	34	33
Limnodynastes convexiusculus	LSS	113	107	*Litoria fallax*	LFX	91	88
Limnodynastes ornatus	LOS	35	66	*Litoria gracilenta*	LGA	22	9
Limnodynastes peronii	LSI	44	43	*Litoria inermis*	LAS	83	74
Limnodynastes tasmaniensis	LSN	50	59	*Litoria latopalmata*	LLA	137	84
Limnodynastes terraereginae	LTE	42	50	*Litoria lesueuri*	LLI	67	205
Mixophyes fasciolatus	MFS	25	31	*Litoria nasuta*	LIA	162	150
Uperoleia fusca	UFA	41	43	*Litoria revelata*	LRT	87	111
Cyclorana alboguttata	CAA	108	111	*Litoria rothii*	LII	3	47
Cyclorana brevipes	CAS	19	14	*Litoria rubella*	LRA	37	37
Cyclorana novaehollandiae	CAE	68	74				

2.1 Datasets

In this study, two frog call datasets, which are collected from subtropical eastern Australia and tropical north-eastern Australia, are used for the experiment. The syllable distribution of both datasets is uneven across 20 frog species. The species name, their acronym, and the number of syllables (described in Section 2.2) are shown in Table 1.

2.2 Syllable segmentation

For frogs, one syllable is an elementary acoustic unit for acoustic classification, which is a continuous frog vocalization emitted from an individual [9, 14]. Here, we apply Härmä's method to perform syllable segmentation for all frog recordings [7]. The distribution of syllable number for all frog species is shown in Table 1, which is highly unbalanced.

2.3 Feature extraction

In this study, we build a feature set consists of nine acoustic features: syllable duration (1), Shannon entropy (1), rênyi entropy (1), tsallis entropy (1), mean frequency (1), average power

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26485v1 | CC BY 4.0 Open Access | rec: 3 Feb 2018, publ: 3 Feb 2018

of frequency band (15), zero crossing rate (1), Mel-frequency cepstral coefficients (19), Linear-frequency cepstral coefficients (19)[1]. All those features have been used for frog call classification except average power of frequency band [8, 14]. All those features are concatenated into together to build a new feature set, whose dimension is 59. To remove the correlation between those features, the normalization is conducted as follows.

$$v_i = \frac{v_i - \mu_i}{\sigma_i} \tag{1}$$

where μ_i and σ_i are the mean and standard deviation computed for each feature i.

2.4 Domain adaptation

In this study, four standard domain adaptation methods are investigated for acoustic frog species classification, which are subspace alignment, transfer component analysis, Geodesic flow kernel, and Information-theoretical learning of discriminant clusters. All those methods do not require labeled target data, which is suitable for our classification task.

2.4.1 Subspace alignment

For subspace alignment (SA), the first step is to generate the subspace for both source and target data [5], which is realized by principal component analysis (PCA). Then, an alignment between those subspaces is learned. The PCA dimensions are determined by minimizing the Bregman divergence between the subspaces.

2.4.2 Transfer component analysis

Transfer component analysis (TCA) is to discover common latent features that share the marginal distribution across the source and target domains while maintaining the intrinsic structure of the original domain data [10]. Specifically, in a reproducing kernel Hilbert space, the latent features are first learned between the source and target domains [12], where the maximum mean discrepancy is used as a marginal distribution measurement criteria [2].

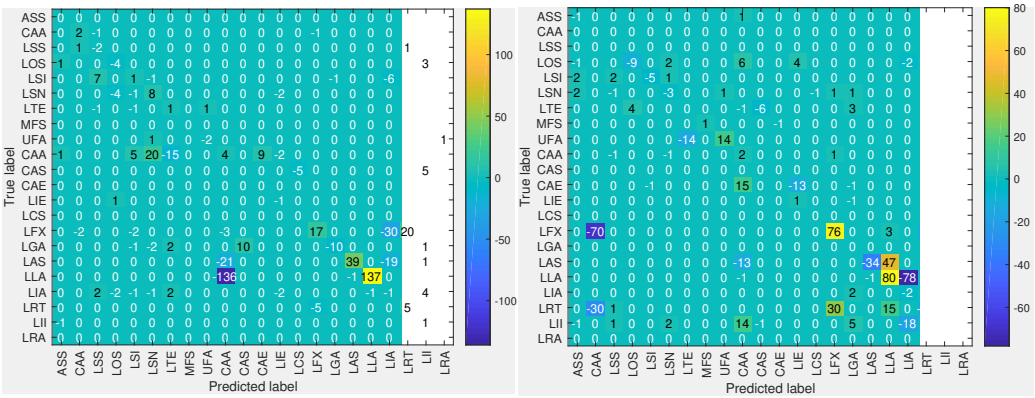
2.4.3 Geodesic flow kernel

Geodesic flow kernel (GFK) aims to find a low-dimensional feature space, which reduces the marginal distribution between labeled source and unlabeled target domains [6]. Specifically, a geodesic flow kernel first projects a large number of subspaces that lie on the geodesic flow curve, which represents incremental differences in geometric and statistical properties between the source and target domain spaces. Then, a classifier is learned from the geodesic flow kernel by selecting features from the geodesic flow curve that are domain invariant.

2.4.4 Information-theoretical learning of discriminant clusters

Information-theoretical learning of discriminant clusters (ITL) learns a domain-invariant feature space and optimizes information-theoretic metrics directly related to discriminative classification on the target domain simultaneously [11]. When ITL identifies a domain-invariant feature space where data in the source and target domains are similarly distributed, the feature space is also learned discriminatingly. This method is realized by optimizing an information-theoretical metric using simple gradient-based methods.

[1]The value in the bracket is the dimension of the feature.



(a) Subtropical East dataset as source data

(b) Tropical North East as source data

Figure 2: Difference between the confusion matrix yielded by the ITL model and the confusion matrix yielded by the no adaptation method.

2.5 Support vector machine

SVMs have been widely used for classifying animal sounds due to their high accuracy and superior generalization properties [1, 9]. To perform the classification, training data is first constructed using the new feature set. Then, the pairs $(v_l^n, L_l^n), l = 1, 2, ..., C_l$ are built using the selected training data, where C_l is the number of frog instance in the training data, v_l^n is the feature vector obtained from the *l-th* frog instance in the training data, L_l^n is the label of frog species n. Finally, the decision function for the classification problem based on SVMs [4] is defined by the training data as follows.

$$f(v) = sgn(\sum_{sv} \alpha_l^n L_l^n K(v, v_l^n) + b_l^n) \tag{2}$$

where $K(.,.)$ is the kernel function, α_l^n is the Lagrange multiplier, and b_l^n is the constant value.

Since our frog call classification task aims to classify multiple classes, one-against-one SVMs is chosen for classifying 22 frog species. To solve this 22 classes pattern recognition problem, 22 classifiers are first constructed and one for each class. Then, the k-th classifier constructs a hyperplane between class n and k-1 classes. A majority vote across all classes is lastly used to give the frog species of the new input instance.

3 Experiments

3.1 Experimental setup

Since the number of frog syllables in both datasets is uneven, the performance is evaluated using a weighted F1-score which is defined as follows:

$$F1\text{-}score = \sum_{i=1}^{n} 2 \cdot \frac{precision(i) \cdot recall(i)}{precision(i) + recall(i)} * r_i \tag{3}$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26485v1 | CC BY 4.0 Open Access | rec: 3 Feb 2018, publ: 3 Feb 2018

where $F1\text{-}score$ denotes the weighted F1-score, $precision$ is defined as $\frac{TP}{TP+FP}$, and $recall$ is defined as $\frac{TP}{TP+FN}$, TP is true positive, TN is true negative, FP is false positive, FN is false negative; i is the maneuver class index of each dataset, r_i is the ratio between the syllable number of one frog species and syllable number of all frog species

The SVMs using an RBF kernel are trained on our proposed feature set of the source domain using LIBSVM [3]. The parameters C and γ were set by searching $2^i(i = -5, -3, -1, 1, 3, 5, 7, 9)$ and $2^j(j = -15, -11, -7, -3, 1, 5)$ to find the optimal value. For SA, the dimensions of the source subspace and target subspace were set at 59. For TCA, we set the subspace bases as 59, and the adaptation regularization parameter λ was set by searching $\lambda \in \{0.01, 0.1, 1, 10, 100\}$. For GFK, the dimensions of the source subspace and target subspace were set at 59. For ITL, the dimensions of the source subspace and target subspace were set at 59, and the optimal regularization parameter λ was set by searching $\lambda \in \{3, 7, 11, 15, 19, 23\}$

3.2 Experimental results and discussions

In this study, we investigate four domain adaptation methods for frog call classification. The classification results are shown in Table 2. We use the fully transductive protocol to evaluate the performances of domain adaptation methods. ITL (Information-theoretical learning of discriminant clusters) achieves the best performance with F1-score of 85.5%. However, the performance of TCA (Transfer component analysis) and GFK (Geodesic flow kernel) is slightly worse than the baseline method. Therefore, it is essential to choose the suitable domain adaptation method for the specific application such as frog call classification.

Among four domain adaptation methods, ITL achieves the best performance. The classification result is sensitive to the parameter λ, which is shown in Fig. 3. For ITL, the best performance for TNE $\rightarrow$ SET and SET $\rightarrow$ TNE is achieved when λ is set at 15 and 11, respectively.

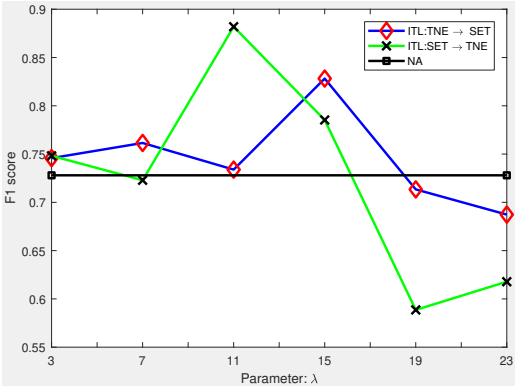


Figure 3: Weighted F1-score on the value of parameter λ.

The differences of confusion matrix between NA and ITL are shown in Fig. 2. Here, along the diagonal, positive value values indicate that the ITL method makes more correct predictions for the corresponding species and negative values indicate the NA method makes more correct predictions for the corresponding species. Off the diagonal, positive values indicate greater confusion by the ITL method for the corresponding pairs of species, while negative values indicate greater confusion by the NA model for the corresponding pairs of species. For TNE

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26485v1 | CC BY 4.0 Open Access | rec: 3 Feb 2018, publ: 3 Feb 2018

Table 2: Weighted classification F1-score of five methods. Here, NA denotes no adaptation baseline. TNE → SET denotes that TNE dataset is used as source data while SET dataset is target data, and vice versa. The reported result is the best performance by optimizing the parameters.

Method	TNE → SET	SET → TNE	Average
NA	72.28%	72.32%	72.80%
SA	72.28%	72.46%	72.80%
TCA	69.58%	68.38%	69.98%
GFK	70.54%	71.49%	71.01%
ITL	**82.82%**	**88.19%**	**85.50%**

→ SET, the performance improvement is due to the improved recognition accuracy of *U.fusa*, *L.fallax* and *L.latopalmata*. For SET → TNE, the performance improvement is due to the accurate recognition of *L.fallax*, *L.inermis* and *L.latopalmata*.

Furthermore, we also test the classification result using the training and testing data from the same area (TNE or SET). We split the data from the same area into 50%-50% for training and testing. The weighted F1-scores for TNE and SET are 98.58% and 97.59%. However, the trained model's generalization is very limited, since training and testing data are collected from the same area. When training and testing data are collected from different areas, the classification performance is decreased to 72.8%. To reduce the impact of mismatch between areas, domain adaptation is used to improve the weighted F1-score from 72.8% to 85.5%. This result verifies the need of domain adaptation for solving the bioacoustic classification problem, especially when large volumes of bioacoustic data are collected from various areas using the distributed sensor network.

4 Conclusions

In this study, we investigate four domain adaptation methods for acoustic frog species classification. First, we segment continuous frog recordings into individual syllables. Then, a novel feature set including syllable duration, Shannon entropy, rênyi entropy, tsallis entropy, fundamental frequency, average power of frequency band, zero crossing rate, Mel-frequency cepstral coefficients, and Linear-frequency cepstral coefficients, is first built. Next, four domain adaptation methods are investigated for discrepancy reduction between source and target domain. Among those domain adaptation methods, ITL provides the best performance, where the weighted classification F1-score can be up to 85.5%.

Although ITL achieves the best performance, linear feature transformation is used. Future work aims to investigate discriminatively learning of nonlinear feature transformation for domain adaptation. In addition, only 22 frog species are used in this paper, a wider variety of frog audio data from different geographical and environmental conditions will be tested in the future experiments.

5 Acknowledgement

The authors would like to thank Mingying Zhu for the help of statistic analysis.

References

[1] Miguel A Acevedo, Carlos J Corrada-Bravo, Héctor Corrada-Bravo, Luis J Villanueva-Rivera, and T Mitchell Aide. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4):206–214, 2009.

[2] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.

[3] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.

[4] Corinna Cortes and Vladimir Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.

[5] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pages 2960–2967, 2013.

[6] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2066–2073. IEEE, 2012.

[7] Aki Harma. Automatic identification of bird species based on sinusoidal modeling of syllables. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 5, pages V–545. IEEE, 2003.

[8] Chenn-Jung Huang, You-Jia Chen, Heng-Ming Chen, Jui-Jiun Jian, Sheng-Chieh Tseng, Yi-Ju Yang, and Po-An Hsu. Intelligent feature extraction and classification of anuran vocalizations. *Applied Soft Computing*, 19(0):1 – 7, 2014.

[9] Chenn-Jung Huang, Yi-Ju Yang, Dian-Xiu Yang, and You-Jia Chen. Frog classification using machine learning techniques. *Expert Systems with Applications*, 36(2):3737–3743, 2009.

[10] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011.

[11] Yuan Shi and Fei Sha. Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. *arXiv preprint arXiv:1206.6438*, 2012.

[12] Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov):67–93, 2001.

[13] Jason Wimmer, Michael Towsey, Paul Roe, and Ian Williamson. Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications*, 23(6):1419–1428, 2013.

[14] Jie Xie, Michael Towsey, Jinglan Zhang, and Paul Roe. Acoustic classification of australian frogs based on enhanced features and machine learning algorithms. *Applied Acoustics*, 113:193–201, 2016.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26485v1 | CC BY 4.0 Open Access | rec: 3 Feb 2018, publ: 3 Feb 2018

[15] Jie Xie, Michael Towsey, Jinglan Zhang, and Paul Roe. Adaptive frequency scaled wavelet packet decomposition for frog call classification. *Ecological Informatics*, 32:134–144, 2016.