Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity

Ilyas Potamitis *

Technological Educational Institute of Crete, Department of Music Technology and Acoustics, E. Daskalaki Perivolia, 74100 Rethymno, Crete, Greece

ARTICLE INFO

Article history:
Received 5 November 2014
Received in revised form 5 January 2015
Accepted 7 January 2015
Available online 15 January 2015

Keywords:
Automatic recognition of species
Large-taxa audio classification
Ecological informatics
Computational ecology

ABSTRACT

A broad range of organisations and individuals are collecting wildlife audio recordings. Huge amounts of audio data have been gathered in the past and since the popularisation of automatic recording units the data are piling up exponentially. The point in gathering them is to analyse them, evaluate insights and hypotheses, identify patterns of activity that are otherwise not apparent and finally design policies on biodiversity issues. For massive volumes of data even visual inspection of spectrograms is unfeasible and interesting cases that could provide valuable insight for concrete hypotheses on the biodiversity status can slip into bliss. In this paper we research a range of techniques that work with minor human supervision. These techniques will construct a dictionary of templates extracted in an unsupervised way from reference recordings and then crawl over a large number of recordings to examine the underlying bioacoustic activity. This work is general and we have applied it to many datasets of animal's vocalisations (e.g. cetaceans, mice, birds). To test our tools objectively and for the sake of reproducibility in this work we report on the MLSP 2013 bird dataset that recently has been publicly released along with all its annotations. We are not interested as to which is the best scoring approach for this dataset. Our aim is to describe novel machine learning tools that try to refine our understanding of biodiversity by answering questions such as: Is the recording under examination void of bird vocalisations or not? If there is bird activity, how many different species are in the recording? What are the most important characteristic spectral segments for recognizing a specific species? The database however is valuable to us to quantify our findings.

1. Introduction

To test the validity of hypotheses concerning issues of biodiversity one needs to collect data over long periods of time and spatial scales (Kelling et al., 2009). In this work we deal with the subset of biodiversity that produces sound. As technology advances, automatic recording units (ARUs) are becoming cheaper, smaller, more power efficient allowing for larger coverage of temporal and spatial scales. Analysis of this data can provide insights about the existence or not of species and their counts can be potentially associated with their threatened status and the general health of the habitat (Zhang et al., 2013). The collection of massive volumes of possibly disparate data requires their organisation, preservation and examination that due to its size requires to be automatic but under the supervision of specialised ornithologists. Machine learning techniques and environmental engineers of various specialisations can work in a complimentary fashion. The data-intensive workflow of biodiversity monitoring starts from the sensors sampling the acoustic space and moves on to the exploratory analysis of this data. New tools are needed to automatically analyse large volumes, confirm or reject hypotheses and summarise results through visualisations (Farina et al., 2011; Pieretti et al., 2011; Sueur et al., 2008; Wimmer et al., 2013). This work is about novel tools that their functionality is assessed on an open database, are unsupervised and therefore can be applied to other datasets if needed with minor human labour and are fast enough to scan large volumes of recordings.

This work builds upon certain publications (Briggs et al., 2012, 2013a, 2013b; Fodor, 2013; Kridler, 2013; Lasseck, 2013, 2014; Potamitis, 2014) and continues their line of thought. It will focus around a core process: The unsupervised extraction of templates from spectrograms and their cataloguing as to become a dictionary of spectral patches belonging to vocalising species. Subsequently, the sweeping of this dictionary through the recordings in order to span the acoustic space of species vocalisations. The collection of templates emerges from the data, as opposed to manually tagging of interesting patterns from the data. We show how the process of cross-correlating the codebook with the recordings returns a single matrix that can be analysed with two different ways: either as a multi-label problem (Briggs et al., 2012, 2013a) or as a regression problem (Fodor, 2013; Lasseck, 2013, 2014). The reason we explore further this school of thought is that the multiple templates approach – in its numerous variations – was a building block for entries that scored first place in five very recent and very competitive bioacoustics recognition

* Tel.: +30 28310 21911.
E-mail address: potamitis@staff.teicrete.gr.

http://dx.doi.org/10.1016/j.ecoinf.2015.01.002
1574-9541/© 2015 Elsevier B.V. All rights reserved.

challenges (see (Bas et al., 2013; Briggs et al., 2013a) and Appendix A), a fact that cannot be disregarded. Therefore, to our point of view, and although it is too early to state this with confidence, it has the potential to become a benchmark for classifying bioacoustics data. The focus of this work is not on best recognition scores as the challenges listed in the Appendix A. Once we clarify the unsupervised dictionary approach originally appearing in Fodor (2013), Lasseck (2013, 2014), and Potamitis (2014) we use it as a vehicle for making several novel tools that we believe are useful as summarising tools. Our novel tools are the following:

a) To be able to report if a recording is void or not of any birds' vocalisation activity (a binary classification problem),
b) To predict the number of different species populating a recording (a regression problem),
c) We will show a principled way on how to find the most informative phrases of a song or call automatically for each species. Interestingly, this is among the standard procedure that ornithologists follow to discern species, especially the dubious cases: they search for a characteristic key of the vocalisation (e.g. the end phrase or on onset or a transition (Marler and Slabbekoorn, 2004, pp. 6–10, p. 149)).
d) We examine the often neglected importance of various forms of metadata, that is, the information from text or sensors other than microphones that come along with a recording and can help classifiers to improve predictions substantially.

All novel tools will be applied to the MLSP 2013 bird dataset of 19 bird species because it is open (https://www.kaggle.com/c/mlsp-2013-birds/data), and the annotation labels of both train and test set have been recently released. Moreover, the predictive platform of Kaggle is still active on this dataset and anyone can reproduce the results of this approach or assess the accuracy of any other by submitting results that will be instantaneously scored. We will also report comments based on various other datasets that are also open to the public and we have analysed in order to support our claims.

This paper is organized as follows: In the Section entitled ‘Methods’ we present: a) the database and its main difficulties as regards species recognition, b) three image analysis techniques that derive spectral templates in an unsupervised way from spectrograms, c) new tools that summarise and visualise bird activity. In the section entitled Results we analyse the fine-tuning of the pattern recognition methods, perform experiments with the database provided and analyse the results of the current work. A discussion of the results concludes this work by presenting possible extensions and summarising the implications of the results.

2. Methods

2.1. The database and its specificities

The MLSP 2013 database is composed of 645 field-recordings each containing vocalisations of 0–5 different species. A subset of 322 recordings is offered as a training set accompanied with known annotation labels from bioacoustics experts and we seek the species vocalising in the other 323 recordings. Labels were provided by confidence-weighted majority voting of a team of experts. Only 192 recordings from the training set are tagged to contain bird vocalisations and the rest are annotated to contain only noise. The hardware used to collect the data was a number of automatic recording units (ARUs) placed at different locations in the H. J. Andrews (HJA) Long-Term Experimental Research Forest, in the Cascade mountain range of Oregon (see also Fig. 5). All recordings are monophonic, sampled at 16,000 Hz with a fixed duration of 10 s each. We did not downsample the original recordings as there are bird vocalisations up to 8 kHz in this dataset. The recordings did not include species vocalising at low-frequencies and much of the lower frequencies were later discarded during the signal processing stage. The training set matches the test set conditions. The list of 19 birds can be found in Table 1.

In Fig. 1 we offer a histogram on the distribution of different number of species to be encountered in all recordings.

Different kinds of difficulties for the recognition task are encountered in the recordings, namely:

a) The dataset includes many faint signals that are hardly perceptible even by visual inspection
b) The dataset is multi-label meaning there is a varying number of species in each recording
c) The dataset is multi-instance meaning there can be a series of vocalisations for each species in the same recording
d) The number of recordings per species is quite small. This has an impact on the machine learning techniques that can be effectively used
e) Heavy wind and rain is present in some recordings.

For various other statistics on the database please refer to (Briggs et al., 2013a).

2.2. Unsupervised construction of a codebook of bird vocalisations

To our knowledge, the first approach of multiple templates on bioacoustic signal classification task was reported in Kridler (2013). In Briggs et al. (2012), the author manually annotates a collection of spectrograms as examples of correct segmentation in order to train a classifier. Both approaches though ground breaking, have a human in the loop as the templates used to scan the data are derived manually by carefully marking the spectrogram and examining the recognition score on a validation dataset in order to decide which templates to keep. The manual touch is a disadvantage: In Kridler (2013) the whole approach is set around a specific problem (i.e. to detect the up-calls of right whales in Challenge no. 1 — see Appendix A). If the same method as set is to be applied to gunshot calls of right whales it will fail as no templates are marked for it. Of course, one could repeat the manual procedure for the new target but this establishes a tight connection between a specialised scientist and the end-user something that the end-user might not want or a later user may not know the assumptions upon which the detector was based. The end-user needs the flexibility of automatic extraction of templates.

In this work as in Potamitis (2014), we are interested in bioacoustics recognition tasks where the spectral blobs of vocalising birds, hereinafter referred to as ‘regions of interest’ (ROIs) are extracted in an

Table 1
List of species in the MLSP 2013 bird recordings dataset.

Class_id	Species
1	Brown creeper
2	Pacific wren
3	Pacific-slope flycatcher
4	Red-breasted nuthatch
5	Dark-eyed junco
6	Olive-sided flycatcher
7	Hermit thrush
8	Chestnut-backed chickadee
9	Varied thrush
10	Hermit warbler
11	Swainson's thrush
12	Hammond's flycatcher
13	Western tanager
14	Black-headed grosbeak
15	Golden crowned kinglet
16	Warbling vireo
17	MacGillivray's warbler
18	Stellar's jay
19	Common nighthawk