Ecological Informatics 44 (2018) 76–81

Contents lists available at ScienceDirect

Ecological Informatics

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

Bird species identification via transfer learning from music genres

Stavros Ntalampiras

Department of Informatics, Università degli studi di Milano, via Comelico 39, 20135 Milan, Italy

A R T I C L E I N F O

Keywords:
Bird species identification
Transfer learning
Echo State Network
Biodiversity monitoring

A B S T R A C T

Humans possess the ability to apply previously acquired knowledge to deal with novel problems quite efficiently. *Transfer Learning* is inspired by exactly that ability and has been proposed to handle cases where the available data come from diverse feature spaces and/or distributions. This paper proposes to transfer knowledge existing in music genre classification to identify bird species, motivated by the existing acoustic similarities. We propose a *Transfer Learning* framework exploiting the probability density distributions of ten different music genres for acquiring a degree of affinity between the bird species and each music genre. To this end, we exploit a feature space transformation based on Echo State Networks. The results reveal a consistent average improvement of 11.2% in the identification accuracy of ten European bird species.

1. Introduction

Monitoring animal communities is becoming increasingly important due to major environmental challenges including invasive species, infectious diseases, climate and land-use change, etc. The effectiveness of the adopted conservation measures is heavily based on the availability of accurate information regarding the range, population size and trends, such that one may assess the conservation status in a species-specific manner. To this end, classical observer-based survey techniques may be employed; nonetheless they are a) costly, b) subject to weather conditions, c) cover a limited amount of time, etc. Thus, in the recent years there has been an increasing interest for automated acoustic monitoring of sound-emitting animals, which may provide reliable information on the presence/absence of target species and on the general biodiversity status of an area.

As a consequence, autonomous recording units (ARUs) have been widely spread among biologists in the recent years to study and analyse different taxonomic groups of sound-producing animals, such as mammals, birds, amphibians, and insects (Fagerlund, 2007; Potamitis, 2015; Grill and Schlüter, 2017; Lopes et al., 2011; Harma, 2003; Fanioudakis and Potamitis, 2017; Potamitis, 2014; Buxton and Jones, 2012; Bardeli et al., 2010; Potamitis et al., 2014) despite their limitations (Wolfgang and Haines, 2016). As ARUs are capable of long-term continuous operation, one is able to collect enormous amount of data in relatively short periods of time. In this context, manual annotation of the acquired audio is not a feasible practise due to cost and time constraints. Thus, automatic analysis of the audio data is mandatory and may facilitate decision making in a series of topics, such as: a) monitoring of range shifts of animal species due to climate change, b) biodiversity assessment and inventorying of an area, c) estimation of species richness and abundance, and d) assessing the status of threatened species.

The basic problems which need to be faced by an automated mechanism dedicated to the classification of bird species are the following: a) it has to operate with reliability across a potentially large number of species (Stowell and Plumbley, 2014), b) it should be able to handle big data as ARUs can capture huge volumes (Aide et al., 2013), and c) be robust to non-stationary environmental noise and sound events (Ntalampiras et al., 2012). As current approaches for large scale bird species identification do not provide high performance, e.g. Wolfgang and Haines (2016) , we propose to complement the way the specific problem is handled so far, and exploit the *Transfer Learning* (TL) logic, i.e. instead of looking at the bird audio signals alone, we propose to statistically analyse them using their similarities with music genres. There are several motivations behind following the specific research path:

- Several musicologists share the belief that the development of music was affected by birdsong to a relatively large extent (Head, 1997; Clark and Rehding, 2005).
- Birds vocalize at traditional scales used in human music (e.g. pentatonic, diatonic) suggesting that birdsong may be thought as music the way humans perceive it (Rothenberg, 2006).
- Famous composers have employed birdsong as a compositional springboard in several genres, e.g. classical (Vivaldi, Beethoven, Wanger, etc.) and jazz (Paul Winter, Jeff Silverbush, etc.) (Franks,

E-mail address: stavros.ntalampiras@unimi.it.
URL: http://sites.google.com/site/stavrosntalampiras/home.

https://doi.org/10.1016/j.ecoinf.2018.01.006
Received 8 September 2017; Received in revised form 25 December 2017; Accepted 31 January 2018
Available online 09 February 2018

2016; Reich, 2010; Thompson, 2014) suggesting perceptual similarities in the respective acoustic structures.

These motivation points suggest that there exists a great variety of perceptual similarities between musical pieces and birdsongs. Thus, tools and methodologies for incorporating knowledge coming from the music information processing field into identifying birdsongs are worth investigating.

This work proposes to exploit the similarities between birdsongs and music genres as assessed by probabilistic modelling for bird species identification. This approach can be grouped under the umbrella of *transfer learning* (Pan and Yang, 2010), which is a relatively new learning framework, where the fundamental characteristic is the transferring of knowledge between two or more classification tasks. Transfer learning has not been extensively exploited by the audio signal processing community. An initial effort among speech and music signals is reported in Coutinho et al. (2014) , where a denoising autoencoder is employed for feature transformation. A more recent method is explained in Phan et al. (2016) where generalized audio events are classified based on their similarities to speech patterns. In the same direction, Lim et al. (2016) train a deep neural network using speech data for sound event classification.

This article aims at presenting the possibilities of transfer learning for bird sound identification. To this end, the dataset is composed of a selection of species covering common European bird species including regular breeding, wintering and migrant ones. On top of that, the Random Forest (RF) classification approach was exploited since it could be considered as a baseline system for many practical applications without making assumptions on the feature set side, such as sequentiality, normalization. More specifically, this work proposes a statistical model approximating the probability density function of each music genre, while taking into account the maximum classification rate criterion. Subsequently, these models are fed with transformed data coming from the birdsong dataset. The probabilities produced by the music genre models are used as features to characterize each species and subsequently classify them using an RF. This process is illustrated in Fig. 1.

2. Transfer learning for bird species identification

This section presents the TL direction followed in this work. We propose to exploit the statistical resemblance between bird calls and musical data as it is assessed via hidden Markov models (HMMs) trained on different music genres, while using the feature space transformation based on Reservoir Networks (*tRN*). The proposed approach encompasses a *tRN* learning a multiple input multiple output transfer function, responsible for transforming features extracted out of bird species to features representative of specific music genres. The following subsections are dedicated to explain the HMM modelling process as well as the training of the Reservoir Network (RN).

2.1. HMM based statistical similarity assessment

The idea behind this approach is that the similarity between music genres and bird species may reveal important information towards the identification of bird species. To this end, we propose to quantify the degree of resemblance between calls coming from a specific species and different music genres using HMMs approximating the distributions exhibited by the music genres. The block diagram of the proposed TL method is shown in Fig. 1.

We assume availability of a music genre dataset $\mathscr{M}$ including N different genres. The mechanism for recognizing music genres relies on well known techniques which have been proven efficient in the last years (Casey et al., 2008). In this work, we are not focussed on the features extraction or the pattern recognition phases; we rather wish to analyse potential transfer of knowledge derived from the music model to the bird call representation. Thus, the feature set comprised the Mel-frequency cepstral coefficients and the classifier is based on left-right HMMs composed of diagonal Gaussian mixture models.

In brief, we employed a triangular Mel scale filterbank for extracting 23 log-energies. Firstly the audio signal is windowed and the short-time Fourier transform (STFT) is computed. The outcome of the STFT passes through the filterbank and the logarithm is computed to adequately space the data. Lastly the discrete cosine transform is applied for decorrelating the data; a procedure which may enhance the performance of the pattern recognition algorithm. The velocity and acceleration coefficients were also appended to the final vector for capturing its dynamics. In the next, the function outputting the features F_v of the audio signal y_t is denoted as f, i.e. $F_v = f(y_t)$.

Pattern recognition is achieved by HMMs, which constitute probabilistic machines able to automatically identify a sequence of patterns within a stream of data (Rabiner, 1989). They are configured in a left-right topology as the nature of the music genres suggests, meaning that only left to right movements are allowed between the states (Ntalampiras, 2014; Panagakis et al., 2010). $\mathscr{M}$ is divided to train T_M and evaluation E_M sets. One HMM is created to represent a specific

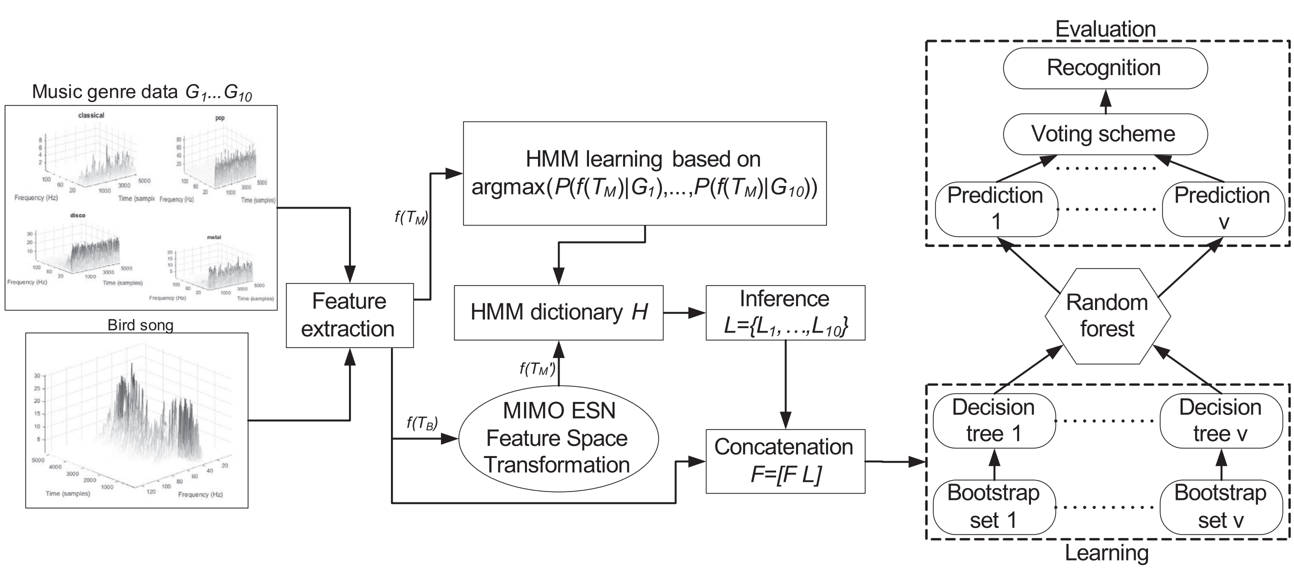

Fig. 1. The block diagram for transfer learning using the statistical affinity based on HMMs trained on music genres (*mHMM*).