Ecological Indicators 80 (2017) 297–302

Contents lists available at ScienceDirect

Ecological Indicators

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

Short Communication

Using non-negative matrix factorisation to facilitate efficient bird species richness surveys

Liang Zhang*, Michael Towsey, Jinglan Zhang, Paul Roe

Queensland University of Technology, Australia

ARTICLE INFO

Keywords:
Non-negative matrix factorisation
Clustering
Acoustic sampling
Bird species richness

ABSTRACT

This paper considers computer-assisted learning of sound spectra in environmental recordings to facilitate manual bird species identification. Today, a variety of automated methods have been successfully applied for acoustic recognition of specific bird species. These methods are more effective for single targeted species detection. For in-field recordings, however, simultaneous vocalisations and unknown species usually make such methods less effective.

In this study, we propose a non-negative matrix factorisation based method to facilitate manual bird species identification from environmental recordings. First, distinct sound spectra are extracted from each audio clip by applying non-negative matrix factorisation and clustering techniques. Based on these distinct sound spectra, a greedy algorithm is then designed to sample audio clips. Each sampled audio clip maximises the number of new spectra. People who follow this sampled sequence of audio clips should be able to identify the most species given a fixed number of audio clips. The efficiency is validated with annotated bird species per minute provided by experienced ornithologists.

1. Introduction

The deployment of acoustic sensors provides a continuous and less invasive approach to record environmental sounds at large spatiotemporal scales. The availability of acoustic data can be used to monitor vocal species such as insects and frogs (Brandes et al., 2006), birds (Acevedo et al., 2009), and bats (Russo and Voigt, 2016). Among these species, birds have been widely recognised as good indicators of biodiversity because they can rapidly reflect environmental changes, they spread over a large landscape, and their ethology is well understood.

A ubiquitous characteristic of bird vocalisations is their diversity. On one hand, inter-specific vocalisations diverge in time and frequency (Michat and Osiejuk, 2010); on the other hand, intra-specific vocalisations vary depending on locations, temperature, or vegetation of a particular landscape (Kosicki and Chylarecki, 2012). A low signal-to-noise ratio of environmental recordings also makes automated detection difficult. In this context, the signal refers to bird vocalisation that is of interest while the noise refers to any unwanted sound such as geophony (rain or wind) and anthropophony (mechanical sounds). Additionally, simultaneous bird vocalisations pose another challenging problem (Briggs et al., 2012).

Despite the acoustic complexity of environmental recordings, human beings are able to differentiate a variety of bird vocalisations by listening to recordings and visually inspecting the spectrograms. Manual analysis can quickly become intractable due to the escalating volume of recordings. An efficient alternative is the use of automated techniques to analyse the recordings. These techniques start with creating statistical models based on the mappings between instances and pre-defined labels. Once the models are created, incoming unlabelled instances can be automatically associated with the pre-defined labels.

Bird species richness is one of the most important studies for biodiversity assessment (Kosicki and Chylarecki, 2014). It is a study of the number of unique bird species in a specific habitat within a specific period of time. We aim with this paper to develop an automated technique to enhance the efficiency of bird species richness surveys with environmental recordings. The problem is formulated as this: given a one-day recording with non-targeted multiple species inventories, identify the maximum number of unique bird species while listening to the minimum number of one-minute audio clips. We specifically focus on audio clips of one day because bird species compositions are relatively stable within this time frame in a specific habitat.

* Corresponding author.
E-mail addresses: l68.zhang@hdr.qut.edu.au (L. Zhang), m.towsey@qut.edu.au (M. Towsey), jinglan.zhang@qut.edu.au (J. Zhang), p.roe@qut.edu.au (P. Roe).

http://dx.doi.org/10.1016/j.ecolind.2017.05.017
Received 25 May 2016; Received in revised form 30 January 2017; Accepted 5 May 2017
Available online 23 May 2017

1.1. Automated bird classification

Classifying bird species by their vocalisations lends itself to rapid analysis of environmental recordings. A typical classification system consists of two primary processes: feature extraction and classification algorithms (Brandes, 2008). Feature extraction is one of, if not the most, crucial steps in such a system. It utilises a value or a vector to represent bird vocalisations in a recording. A prevalent form of feature extraction is the spectrogram, which is a result of using the short-time Fourier transform to convert a waveform into multiple spectra. Mel-frequency cepstral coefficients (Kogan and Margoliash, 1998) is one of the most commonly used approaches to summarise bird vocalisations from spectrograms. It is developed to capture human speech, but is not necessarily suitable for bird vocalisations. Other feature extraction methods include sinusoidal pulses with time-varying amplitude and frequency (Harma, 2003), spectral peak tracks (Chen and Maher, 2006), and syllable pair histograms (Somervuo and Harma, 2004). They have been used individually or in combination for bird vocalisation representations. These features do well in capturing specific types of bird vocalisations; however, to achieve this goal, people should know in advance what types of bird vocalisations are in the recordings.

The classification algorithms are about using statistical criteria to map extracted features with some pre-defined labels. Multivariate analysis (Martindale, 1980) and cross correlation (Clark et al., 1987) are the simplest two algorithms to match similar bird vocalisation with templates, but they are prone to errors. More advanced algorithms include artificial neural network (McIlraith and Card, 1997), hidden Markov models (Kogan and Margoliash, 1998), decision tree (Vilches et al., 2006), and support vector machine (Fagerlund, 2007). These algorithms have been successfully applied to deal with specific sets of bird vocalisations. Multiple simultaneous bird vocalisations remain a difficult problem for automated species identification. Recently, a multi-instance multi-label classification method has been proposed to tackle the problem of simultaneous vocalisations in environmental recordings (Briggs et al., 2012). This method is a supervised method that requires massive training data.

Automated classification techniques offer a promising approach for bird species analysis, especially when people are confronted with a large number of recordings. However, acquiring a labelled dataset for statistical model generation is sometimes laborious and expensive. For bird species richness surveys, one might not be able to know what species are in the recordings beforehand. This paper aims to develop an approach to ameliorate such difficulty by automatically extracting distinct vocalisations from audio recordings. These vocalisations can further be used as a proxy to direct bird species richness surveys.

1.2. Non-negative matrix factorisation

Non-negative matrix factorisation (Lee and Seung, 1999) can decompose a matrix into a product of two matrices. The felicity of such decomposition is it enables to generate a parts-based representation, enabling to characterise distinct bird vocalisations of a spectrogram automatically. Since its inception, non-negative matrix factorisation has seen a broad range of applications, including multiple sound sources separation (Smaragdis, 2004; Zhang et al., 2008), music transcription (Bertin et al., 2007), and gene data expression (Frigyesi and Höglund, 2008; Hutchins et al., 2008). Recently, probabilistic latent component analysis (PLCA) – a probabilistic variant of non-negative matrix factorisation has been proposed for the analysis of soundscape ecology (Eldridge et al., 2016). This paper also motivates our work.

Non-negative matrix factorisation is described as follows. Given a matrix ***S*** of size $n \times m$, it can be represented by the multiplication of two non-negative matrices ***W*** and ***H***:

$$\boldsymbol{S} \approx \boldsymbol{W}\cdot\boldsymbol{H} \tag{1}$$

where the matrix ***W*** has a size of $n \times r$ and the matrix ***H*** has a size of $r \times m$. Here, r is called the *factorisation rank*, affecting the performance of the approximation.

The approximation is achieved by minimising a cost function that measures approximation error. One of the cost functions is:

$$D = \frac{\|\boldsymbol{S} - \boldsymbol{W}\cdot\boldsymbol{H}\|_{\mathrm{F}}}{\sqrt{n \times m}} \tag{2}$$

where D is the *root-mean-squared (RMS) residual*. The subscript ‘F’ denotes the *Frobenius* norm. Let a_{ij} be an element of matrix $\boldsymbol{S}–\boldsymbol{W}\cdots\boldsymbol{H}$, the *Frobenius* norm is calculated as:

$$\|\boldsymbol{S} - \boldsymbol{W}\cdot\boldsymbol{H}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} |a_{ij}|^2} \tag{3}$$

The algorithm is iterative starting with random initial values for matrices ***W*** and ***H***. During the iteration, the matrices ***W*** and ***H*** are updated using the following equations:

$$\boldsymbol{W}_{iq} \longleftarrow \boldsymbol{W}_{iq}\cdot\frac{(\boldsymbol{S}\cdot\boldsymbol{H}^T)_{iq}}{(\boldsymbol{W}\cdot(\boldsymbol{H}\cdot\boldsymbol{H}^T))_{iq}} \quad \text{for } 1 \le i \le n \text{ and } 1 \le q \le r \tag{4}$$

$$\boldsymbol{H}_{qj} \longleftarrow \boldsymbol{H}_{qj}\cdot\frac{(\boldsymbol{W}^T\cdot\boldsymbol{S})_{qj}}{((\boldsymbol{W}^T\cdot\boldsymbol{W})\cdot\boldsymbol{H})_{qj}} \quad \text{for } 1 \le q \le r \text{ and } 1 \le j \le m \tag{5}$$

The key parameter in Eqs. (4) and (5) is factorisation rank r, which determines the size of matrices ***W*** and ***H***. Other variants of non-negative matrix factorisation algorithm differ in the non-negativity constraints on the matrix ***W***, the matrix ***H***, or both (Hoyer, 2004; Pascual-Montano et al., 2006; Tao et al., 2002). A simple example of non-negative matrix factorisation on a spectrogram can be found in this paper (Smaragdis, 2004). Generally, the columns of the matrix ***W*** denote the distinct spectral profiles and the rows of the matrix ***H*** denote the corresponding temporal coefficients of each spectral profile.

This paper aims to develop a sampling technique to facilitate manual bird species identification in environmental audio recordings. The difficulty in developing a sampling technique lies in the accurate detection of bird vocalisations by computers. However, most approaches require prior knowledge of different types of bird vocalisations, which is not practical when the size of recording is large. Non-negative matrix factorisation offers a potential solution to such a problem.

2. Methods

The general process of our method is described as follows. First, we apply the non-negative matrix factorisation to decompose spectrograms into spectral profile matrices and temporal coefficient matrices. A hierarchical clustering technique is then used to generate distinct spectra of bird vocalisations from the decomposed matrices. Finally, audio clips are sampled in a sequence with the maximum number of new distinct spectral profiles. Following this sampled sequence of audio clips, people should be able to find the most bird species given a fixed number of audio clips.

2.1. Estimating the factorisation rank

The most crucial issue in this study is to determine a proper factorisation rank r for the non-negative matrix factorisation. There is no uniform r for the non-negative matrix factorisation due to the inherent complexity of environmental recordings. A common solution is to optimise the factorisation performance by increasing r (Brunet et al., 2004; Hutchins et al., 2008).

This work follows an adaptive method proposed by Frigyesi and Höglund (2008) to determine the factorisation rank r based on the acoustic complexity of each recording. For two consecutive factorisation rank $r - 1$ and r, we calculate the decreases of root-mean-squared (RMS) residual for the original spectrogram (ΔD_{o}) and its randomised