Applied Soft Computing 19 (2014) 1–7

Contents lists available at ScienceDirect

Applied Soft Computing

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

Intelligent feature extraction and classification of anuran vocalizations

Chenn-Jung Huang[a,b,*], You-Jia Chen[b], Heng-Ming Chen[b], Jui-Jiun Jian[c], Sheng-Chieh Tseng[a], Yi-Ju Yang[d], Po-An Hsu[e]

[a] Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan
[b] Department of Electrical Engineering, National Dong Hwa University, Taiwan
[c] Institute of Electrical Engineering, National Dong Hwa University, Taiwan
[d] Institute of Ecology and Environmental Education, National Dong Hwa University, Taiwan
[e] Institute of Computer Science, National Tsing Hua University, Taiwan

ARTICLE INFO

Article history:
Received 8 July 2010
Received in revised form 10 March 2012
Accepted 18 January 2014
Available online 5 February 2014

Keywords:
Data mining
Clustering analysis
Neural networks
Feature selection
Segmentation

ABSTRACT

An intelligent identification system for mixed anuran vocalizations is developed in this work to provide the public to easily consult online. The raw mixed anuran vocalization samples are first filtered by noise removal, high frequency compensation, and discrete wavelet transform techniques in order. An adaptive end-point detection segmentation algorithm is proposed to effectively separate the individual syllables from the noise. Six features, including spectral centroid, signal bandwidth, spectral roll-off, threshold-crossing rate, spectral flatness, and average energy, are extracted and served as the input parameters of the classifier. Meanwhile, a decision tree is constructed based on several parameters obtained during sample collection in order to narrow the scope of identification targets. Then fast learning neural-networks are employed to classify the anuran species based on feature set chosen by wrapper feature selection method. A series of experiments were conducted to measure the outcome performance of the proposed work. Experimental results exhibit that the recognition rate of the proposed identification system can achieve up to 93.4%. The effectiveness of the proposed identification system for anuran vocalizations is thus verified.

1. Introduction

Pattern recognition forms a fundamental solution to different problems in real world applications [1]. The function of pattern recognition is to categorize an unknown pattern into a distinct class based on a suitable similarity measure. Thus similar patterns are assigned to the same classes while dissimilar patterns are classified into different classes.

In speech recognition, a source model is assumed and the signal is expected to obey the laws of a specific spoken language with vocabulary and grammar. Anuran vocalization is a representative instance of a category of natural sounds where a vocabulary and other structural elements are expected. In comparison with the human speech recognition problem, animal sounds are usually simpler to recognize. Speech recognition often proceeds in a quiet and similar environment, while anuran's sounds are usually recorded in a much noisier environment, under which we must recognize simpler vocalizations.

In this work, an automatic identification system for mixed anuran vocalizations is proposed to recognize the anuran species based on the recorded audio signals that were sampled from recordings of anuran sounds in an outdoor environment. The sampled signals were first converted into frequency signals. Then syllable segmentation and feature selection methods are employed to separate the original anuran calls into syllables and to derive the input feature sets for the classifiers. Experimental results reveal that the proposed identification system is effective in identifying mixed anuran calls.

There are several major contributions in this research. Firstly, to the best of our knowledge, this work is the first research that addresses the problem on identification of mixed anuran vocalizations in the literature. Secondly, surrounding noise has always been ignored in researches about animal sound recognition system such as those of literature [5] and is now considered in the present research. A novel segmentation method is developed in order to tackle the noise problem. The maximum amplitude of vibration is selected to successfully segment the syllables. Thirdly, instead of human voice samples, the collection of anuran vocalization samples is difficult stemmed from several factors, such as the presence of noise, anuran seasonal habitual behavior and nighttime calls. This work adopts the clustering analysis before the classification stage to reduce the complexity of classification and effectively increase

* Corresponding author at: Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan. Tel.: +886 912796179.
E-mail addresses: cjhuang@mail.ndhu.edu.tw, chennjunghuang@gmail.com (C.-J. Huang).

1568-4946/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2014.01.030

the recognition rate. The last but not the least, the algorithm proposed in this work is not only specifically designed for detecting the sounds of frog species in Taiwan, but also provides the public a friendly education gateway for recognizing the frog species in Taiwan.

The remainder of this article is organized as follows. In Section 2, we describe related work in audio recognition. The architecture of the intelligent identification system for anuran vocalizations is presented in Section 3. Experimental results and analysis are given in Section 4. Finally, Section 5 concludes the work.

2. Related work

Features used in sound recognition applications are usually chosen such that they represent some meaningful characteristics. The selection of actual features used in recognition is a critical part for the recognition system. Anuran sounds can be taken as an organized sequence of brief sounds from a species-specific vocabulary. Those brief sounds are usually called syllables [2]. Through the use of preprocessing, we can extract those useful syllables and separate the anuran calls from non-anuran sounds.

Recently, most of the researches on sound recognition for animal calls have been focusing on animal species identification, such as that of bird species [3,5]. Identification of different species of animals based on the recorded calls helped people understanding animal calls. Tyagi et al. [3] introduced a new representation for bird syllables which was based on the long-term average spectrum and template matching was adopted in classification tasks. Vilches et al. [4] used data mining techniques for classification and analyses were performed on a pulse-by-pulse basis in contrast to traditional syllable-based systems. Somervuo et al. [5] studied different parametric representations of bird syllables. Notably, all of the above-mentioned research work focused on bird calls. The collection of bird's call samples turn out to be not so difficult because most of the birds can be seen and their calls can be heard during the daytime.

In the applications of sound recognition, the call boundary detection is an essential problem to be resolved. To a great extent, the performance of sound recognition depends deeply on whether call boundary detection algorithm can perfectly detect the end points of the syllables. This construct of segmentation before recognition process assists the system in finding meaningful and complete calls. Proper segmentation of anuran calls can exhibit the characteristic of the calls, thereby leading to higher accuracy in anuran recognition.

Some investigations focused on recognition of different animal calls [6,7]. Mitrovic et al. [6] used machine learning technology to recognize different animal calls, including those of birds, cats, cows and dogs. Guodong et al. [7] employed support vector machines to identify 16 different classes of animal sounds. In this work, the non-anuran sound detection is different from the above approaches because it is impossible to obtain the samples that are categorized as non-anuran species. Although outlier detection in large data sets is an active research field in data mining and there are numerous application domains that can lead to illegal or abnormal behavior, such as fraud detection [8], network intrusion detection, insurance fraud, medical diagnosis, marketing, or customer segmentation, the resolution of the type of outlier detection problems, such as the non-anuran species detection problem tackled in this work, to the best our knowledge, is never investigated in the literature.

3. Architecture of the identification system for mixed anuran vocalizations

The architecture of the identification system for mixed anuran vocalizations as shown in Fig. 1 can be divided into five main modules, including signal preprocessing, syllable segmentation, feature extraction, clustering analysis, and classification modules. Undesirable information is first removed from the raw input signals in order to preserve the desired characteristics of anuran calls during the signal preprocessing stage. The resulting signal is then segmented by the syllable segmentation method and the segmented syllables are further processed with the feature extraction algorithms. A well-known feature selection method, wrapper approach, is employed to find the best combinations of the feature sets that are suitable to the classifier used in the present research. The species of the mixed anuran calls are then determined by the classifier.

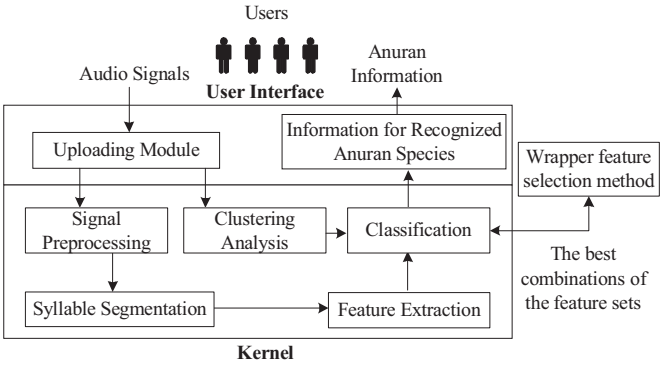

Fig. 1. Architecture of identification system for mixed anuran vocalizations.

3.1. Signal preprocessing

The recorded sound signal is resampled at 16 kHz frequency and saved as 16-bit mono format. The amplitude of each sound signal is normalized within the range [−1, 1] for the ease of further processing. Pre-emphasis and de-noise are applied in order to "purify" the data in the noisy environment.

The motivation of using pre-emphasis technique is to compensate the high-frequency part that is suppressed during the recording of the audio signals by using sound production mechanism. De-noise filter is employed to remove the noise during the signal analysis.

3.1.1. Pre-emphasis filter

Pre-emphasis filter is used to amplify the importance of high-frequency formants in this work. Base on the input sequence, the signal derived by pre-emphasis technique can be expressed by,

$$S'_n = S_n - \alpha \times S_{n-1}, \quad (1)$$

where S_n and S_{n-1} denote the raw signal sequence, and α is obtained from $H(z) = 1 - \alpha \times z^{-1}$. Here the z-transform $H(z)$ represents a high-pass filter.

3.1.2. De-noise filter

De-noise filter is a well-known technique to remove the noise during the signal analysis. De-noise filter kernel function makes use of wavelet threshold function in 1-dimension signal. The output signal of De-noise filter is expressed by,

$$D'_n = f(g(h(D_n, c_1, c_2))), \quad (2)$$

where D_n denotes the raw signal, c_1 and c_2 represent the coarsest level and quadrature mirror filter, respectively, and h, g and f stand for the forward wavelet transform, soft threshold function, and the inverse wavelet transform, respectively.