Expert Systems with Applications 39 (2012) 9986–9993

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Serial combination of multiple classifiers for automatic blue whale calls recognition

Mohammed Bahoura [a,*], Yvan Simard [b,c]

[a] Department of Engineering, Université du Québec à Rimouski, 300, allée des Ursulines, Rimouski, Qc, Canada G5L 3A1
[b] Marine Sciences Institute, Université du Québec à Rimouski, 310, allée des Ursulines, Rimouski, Qc, Canada G5L 3A1
[c] Maurice Lamontagne Institute, Fisheries and Oceans Canada, 850 Route de la Mer, P.O. Box 1000, Mont-Joli, Qc, Canada G5H 3Z4

ARTICLE INFO

Keywords:
Blue whale calls
Feature extraction
Chirplet transform
Short-time Fourier transform
Neural network
Vector quantization
k-Nearest neighbour
Classifiers combination

ABSTRACT

In this paper, we propose a serial combination architecture of classifiers for automatic blue whale calls recognition. Based on class's best selection operator, the proposed system uses a best classifier for D call class followed by another one that efficiently discriminate the A and B calls. The first classifier uses the short-time Fourier transform to characterize the patterns, while the second uses the chirplet transform. Both classifiers are based on multi-layer perceptron neural network. The classification performance (95.55%) of the proposed system outperforms all tested single classifiers. The other advantages of the system are no requirement for adjusting a series of parameters and simple feature extraction.

1. Introduction

Multiple classifier combination has become a very active research area in the last two decades. These activities are motivated by the expectation that classification errors can be reduced if an ensemble of classifiers rather a single classifier is used for a given task (Last, Bunke, & Kandel, 2002). Thus, the classifier combination has been successfully applied to handwritten character recognition (Xu, Krzyzak, & Suen, 1992), speaker recognition (Altınçay & Demirekler, 2003; Chen, Wang, & Chi, 1997), face identification (Czyz, Kittler, & Vandendorpe, 2004), biomedical engineering (Guler & Ubeyli, 2005), financial distress prediction (Sun & Li, 2009), etc. There are two ways of combining classifiers: parallel or serial combination (Figs. 1 and 2). Parallel combination arranges classifiers in parallel and the results obtained concurrently are integrated by a combining algorithm. Serial combination sequentially applies classifiers and the unclassified pattern from the prior is fed to the next classifier (Kim, Kim, & Lee, 2002). The performance of the parallel structure depends on the combination algorithm. The commonly used algorithms include voting principle, evidential reasoning, Bayesian formalism, Dempster–Shafer theory, etc. (Chen et al., 1997). In serial structure the order of arrangement is crucial for the classification performance of the system. Neagu, Guo, and Wang (2006) proposed a serial combination mechanism based on the class-wise expertise of diverse classifiers, which was recently modified by Sun and Li (2009). As shown in Fig. 3, this serial combination architecture consists of each class's best classifier and the wholly best classifier (Sun & Li, 2009).

The rest of the paper is organized as follows: Section 2 describes the test database. Section 3 presents the feature extraction methods and the classification techniques that are combined to construct the individual classifiers. Section 4 describes the proposed serial combination method of classifiers. Section 5 is about experimental results, in which serial combination system is constructed and evaluated on the test database. Finally, conclusion is given in Section 6.

2. Database

2.1. Collection of data

The recordings were collected in the Saguenay-St. Lawrence Marine Park, in the blue whale feeding ground at the head of the Laurentian channel in the Lower St. Lawrence Estuary. The hydrophones were AURAL autonomous hydrophones (Multi-Electronique Inc., Rimouski, Qc, Canada) moored at intermediate depths in the 300-m high water column where a well defined sound channel develops in summer. The 16-bit wave recording rate was set at 2000 Hz, which includes an appropriate antialiasing low-pass filter (1000 Hz). Details can be found in (Simard & Roy, 2008). The calls of interest here are the signature calls of North Atlantic blue whales; notably the A and B infrasounds (15–20 Hz), often occurring together in AB phrases, and the audible D-call (35–120 Hz) also known as arch sound (Berchok, Bradley, & Gabrielson, 2006). The D call is more variable than A and B calls (Mouy, Bahoura, & Simard, 2009). Details on the St. Lawrence blue wall call repertoire can be found in (Berchok et al., 2006). As shown in the examples of Fig. 4, the recordings in the call bands are usually corrupted with various kinds of noises.

* Corresponding author. Tel.: +1 418 723 1986.
E-mail address: Mohammed_Bahoura@uqar.qc.ca (M. Bahoura).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.156

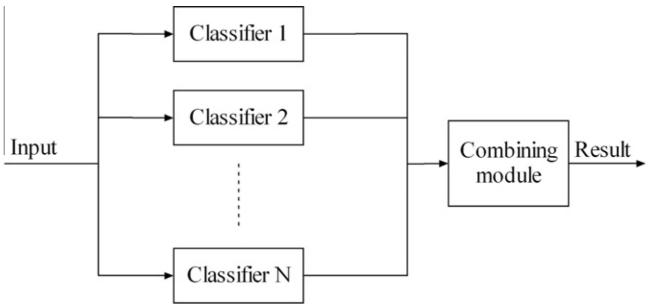

Fig. 1. Parallel combination of multiple classifiers (Kim et al., 2002).

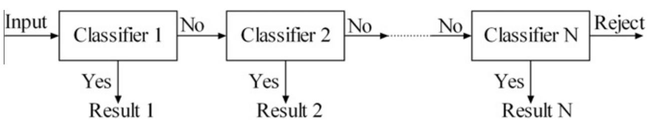

Fig. 2. Serial combination of multiple classifiers (Kim et al., 2002).

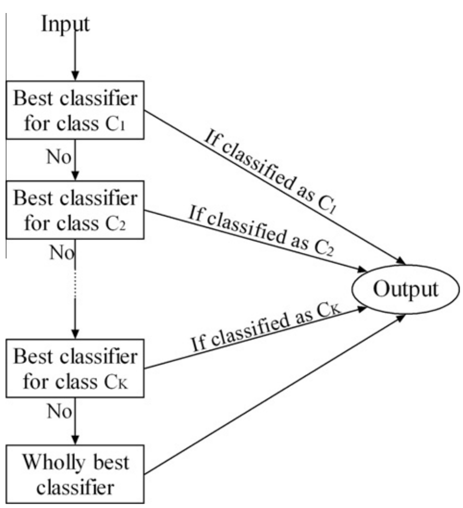

Fig. 3. Serial combination of multiple classifiers based on the class-wise expertise (Neagu et al., 2006).

2.2. Data preprocessing

Blue whale vocalizations were extracted manually and categorized into the three classes (A, B and D) by visualizing their spectrogramme using Adobe Audition (www.adobe.com). Each class of the test database contains 100 calls. Due to the limited size of this test dataset, the *"k-fold cross-validation"* method is employed to evaluate the performance of classifiers. Each class was divided into 10 groups and each test randomly used 9 groups for training the other one for testing.

3. Single classifiers

Typical classification systems include two blocks: feature extraction and modeling/classification (Bahoura, 2009). They generally operate in two steps: training and testing. During the training step, each class of data is modeled to determinate a discriminant delimiting different classes. During the testing step, the new data are classified using the discriminant.

In this section, we describe three feature extraction methods and three classification techniques that are used to construct individual classifiers. These classifiers are later evaluated to select the ones that can build the best multi-classifier serial combination.

3.1. Feature extraction

In this section we present three methods that have recently been proposed to characterize blue whale calls (Bahoura & Simard, 2008, 2010). Feature extraction methods based on short-time Fourier transform (STFT) and wavelet packet transform (WPT) were successfully applied to classify Blue whale calls into A, B and D classes (Bahoura & Simard, 2010). Characterization by chirplet transform was shown to efficiently discriminate the A and B calls (Bahoura & Simard, 2008). More details can be found in (Bahoura & Simard, 2008, 2010).

3.1.1. Fourier transform

The short-time Fourier transform (STFT) of a discrete-time signal $s[n]$ is a Fourier transform performed in successive frames:

$$S[m,k] = \sum_{n=0}^{N-1} s[n]w[n-mL]e^{-j2\pi nk/N} \tag{1}$$

where $w[n]$ is a short-time windowing function of size L, centered at time location m and N is the number of discrete frequencies ($N \geqslant L$). In this work, we used a Hamming window function of length N = 512 and a 50% overlap. The power spectrum density (PSD) is given by:

$$P_s[m,k] = \frac{1}{N}|S[m,k]|^2 \tag{2}$$

The STFT-based characterization method extracted features from two subbands, (15.625–20.996 Hz) and (39.062–85.449 Hz), respectively corresponding to the AB and D calls frequency ranges. For a sampling frequency (f_S = 250 Hz), the first six components of the feature vector $\mathbf{x}_m$ were obtained by sequentially averaging PSD points between $P_s[m,31]$ and $P_s[m,42]$ by bins of 2 points. The last six features were similarly obtained for the PSD interval from $P_s[m,79]$ to $P_s[m,174]$ but with bins of 16 points. For a given frame m, the feature vector component $x_m[n]$ is defined by:

$$\tilde{x}_m[n] = \begin{cases} \frac{1}{2}\sum\limits_{k=29+2n}^{30+2n} P_s[m,k] & n = 1,2,\ldots,6. \\ \frac{1}{16}\sum\limits_{k=63+16(n-6)}^{78+16(n-6)} P_s[m,k] & n = 7,8,\ldots,12 \end{cases} \tag{3}$$

Hence, a 12-dimensional feature vector $\mathbf{x}_m = [\tilde{x}_{m,1}, \tilde{x}_{m,2}, \ldots, \tilde{x}_{m,12}]^T$ is constructed to classify an unknown call into A, B or D classes, where T represents the transpose operation. However, only a 6-dimensional feature vector $\mathbf{x}_m = [\tilde{x}_{m,1}, \tilde{x}_{m,2}, \ldots, \tilde{x}_{m,6}]^T$ is constructed for its classification into A or B classes.

3.1.2. Wavelet packet transform

For a given level j, the WPT decomposes the input signal $s[n]$ of length N into 2^j subbands corresponding to the set of wavelet coefficients.

$$w_k^j[n] = \mathrm{WPT}\{s[n], j\} \tag{4}$$

where $w_k^j[n]$ defines the nth coefficient of the kth subband, where $n = 0, \ldots, \frac{N}{2^j} - 1$ and $k = 0, \ldots, 2^j - 1$. In fact, n is the time index and k is the frequency index. For a given sampling frequency (f_S = 250 Hz), the time and frequency resolutions depend on the frame length N and the level j. For comparing this characterization method with the previous one, the WPT is computed for $j = 7$ using rectangular window function of length $N = 512$ with a 50% overlap.