Expert Systems with Applications 38 (2011) 620–630

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Text-independent speaker verification using ant colony optimization-based selected features

Shahla Nemati [a,1], Mohammad Ehsan Basiri [b,*]

[a] Islamic Azad University, Arsanjan Branch, Fars, Iran
[b] Computer Engineering Department, University of Isfahan, Hezar Jerib Avenue, Isfahan, Iran

ARTICLE INFO

Keywords:
Speaker verification
Gaussian mixture model universal background model (GMM-UBM)
Feature selection
Ant colony optimization (ACO)
Genetic algorithm (GA)

ABSTRACT

With the growing trend toward remote security verification procedures for telephone banking, biometric security measures and similar applications, automatic speaker verification (ASV) has received a lot of attention in recent years. The complexity of ASV system and its verification time depends on the number of feature vectors, their dimensionality, the complexity of the speaker models and the number of speakers. In this paper, we concentrate on optimizing dimensionality of feature space by selecting relevant features. At present there are several methods for feature selection in ASV systems. To improve performance of ASV system we present another method that is based on ant colony optimization (ACO) algorithm. After feature reduction phase, feature vectors are applied to a Gaussian mixture model universal background model (GMM-UBM) which is a text-independent speaker verification model. The performance of proposed algorithm is compared to the performance of genetic algorithm on the task of feature selection in TIMIT corpora. The results of experiments indicate that with the optimized feature set, the performance of the ASV system is improved. Moreover, the speed of verification is significantly increased since by use of ACO, number of features is reduced over 80% which consequently decrease the complexity of our ASV system.

1. Introduction

Automatic speaker recognition (ASR) systems are generally divided into two categories, namely: automatic speaker identification (ASI) systems which are designed to answer the question “who is the speaker?” or automatic speaker verification (ASV) systems that aim to answer the question “is the speaker who they claim to be?”.

ASV refers to the task of verifying speaker’s identity using speaker-specific information contained in speech signal. Speaker verification methods are totally divided into text-dependent and text-independent applications. When the same text is used for both training and testing, the system is called to be text-dependent while for text-independent operation, the text used to train and test of the ASV system is completely unconstrained. Text-independent speaker verification requires no restriction on the type of input speech. In contrast, Text-independent speaker verification usually gives less performance than text-dependent speaker verification, which requires test input to be the same sentence as training data (Xiang & Berger, 2003).

Applications of speaker verification can be found in biometric person authentication such as an additional identity check during credit card payments over the Internet while, the potential applications of speaker identification can be found in multi-user systems. For instance, in speaker tracking the task is to locate the segments of given speaker(s) in an audio stream (Kwon & Narayanan, 2002; Lapidot, Guterman, & Cohen, 2002; Liu & Kubala, 1999; Martin & Przybocki, 2001). It has also potential applications in automatic segmentation of teleconferences and helping in the transcription of courtroom discussion.

Speech signals contain a huge amount of information and can be described as having a number of different levels of information. At the top level, we have lexical and syntactic features, below that are prosodic features, further below these are phonetic features, and at the most basic level we have low-level acoustic features, which generally give information on the system that creates the sound, such as the speakers’ vocal tract. Information solely about how the sound is produced (from low-level acoustic features) should give enough information to identify accurately a speaker, as this is naturally speaker dependent and independent of text (Day & Nandi, 2007).

Low-level acoustic features also contain some redundant features, which can be eliminated using feature selection (FS) techniques. The objective of FS is to simplify a dataset by reducing its

* Corresponding author. Tel.: +98 3117932671; fax: +98 3117932670.
E-mail addresses: nemati@iaua.ac.ir (S. Nemati), basiri@eng.ui.ac.ir (M.E. Basiri).
1 Tel.: +98 7297623908; fax: +98 7297624806.

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2010.07.011

dimensionality and identifying relevant underlying features without sacrificing predictive accuracy. By doing that, it also reduces redundancy in the information provided by the selected features (Jensen, 2005). In real world problems, FS is a must due to the abundance of noisy, irrelevant or misleading features. Selected features should have high inter-class variance and low intra-class variability. Ideally, they should also be as independent of each other as possible in order to minimize redundancy.

Feature selection is extensive and it spreads throughout many fields, including signal processing (Nemati, Boostani, & Jazi, 2008), face recognition (Kanan, Faez, & Hosseinzadeh, 2007), text categorization (Aghdam, Ghasem-Aghaee, & Basiri, 2009), data mining and pattern recognition (Jensen, 2005; Kanan et al., 2007). FS has been rarely used in ASV systems. Day and Nandi (2007) employed genetic programming (GP) for FS; also Pandit and Kittkr (1998) use L plus–R minus feature selection algorithm for text-dependent speaker verification. Cohen and Zigel (2002) employed dynamic programming for FS in speaker verification and Ganchev, Zervas, Fakotakis, and Kokkinakis (2006) used information gain (IG) and gain ratio (GR) for FS in ASV task.

Among too many methods that are proposed for FS, population-based optimization algorithms such as genetic algorithm (GA)-based method and ant colony optimization (ACO)-based method have attracted a lot of attention. These methods attempt to achieve better solutions by application of knowledge from previous iterations.

Genetic algorithms are optimization techniques based on the mechanism of natural selection. They used operations found in natural genetics to guide itself through the paths in the search space (Siedlecki & Sklansky, 1989). Because of their advantages, recently, GAs have been widely used as a tool for feature selection in data mining (Srinivas & Patnik, 1994).

Dorigo and Caro (1999) represented meta-heuristic optimization algorithm based on behavior of ants in the early 1990s. ACO is a branch of newly developed form of artificial intelligence called Swarm Intelligence (SI). Formally, swarm intelligence refers to the problem-solving behavior that emerges from the interaction of cooperative agents, and computational Swarm Intelligence (CSI) refers to algorithmic models of such behavior. ACO algorithm is inspired by social behavior of ant colonies. Although they have no sight, ants are capable of finding the shortest route between a food source and their nest by chemical materials called pheromone that they leave when moving (Liu, Abbass, & McKay, 2004).

ACO algorithm was firstly used for solving traveling salesman problem (TSP) (Dorigo, Maniezzo, & Colorni, 1996) and then has been successfully applied to a large number of difficult problems like the quadratic assignment problem (QAP) (Maniezzo & Colorni, 1999), routing in telecommunication networks, graph coloring problems, scheduling, etc. This method is particularly attractive for feature selection, as there seems to be no heuristic that can guide search to the optimal minimal subset every time (Kanan et al., 2007).

In our previous work, we have proposed an ACO algorithm for feature selection in GMM-based ASV systems (Nemati et al., 2008). In this paper, we propose some modifications to the algorithm and apply it to larger feature vectors containing mel-frequency cepstral coefficients (MFCCs) and their delta coefficients, two energies, Linear prediction cepstral coefficients (LPCCs) and their delta coefficients. Then, feature vectors are applied to a Gaussian mixture model universal background model (GMM-UBM) which is a text-independent speaker verification model. Finally, the classifier performance and the length of selected feature vector are considered for performance evaluation.

The rest of this paper is organized as follows. Section 2 presents a brief overview of ASV systems. Feature selection methods are described in Section 3. Ant colony optimization is described in Section 4. Section 5 explains the proposed feature selection algorithm. Genetic algorithms are described in Section 6. Section 7 reports computational experiments. It also includes a brief discussion of the results obtained and finally the conclusion and future works are offered in the last section.

2. An overview of ASV systems

The typical process in most proposed ASV systems involves some form of preprocessing of the data (silence removal) and feature extraction, followed by some form of speaker modeling to estimate class dependent feature distributions (see Fig. 1). A comprehensive overview can be found in Campbell (1997). Adopting this strategy the ASV problem can be further divided into the two problem domains of:

(1) Preprocessing, feature generation and selection.
(2) Speaker modeling and matching.

These steps are described in following sections in more details.

2.1. Feature extraction

The original signal, the speech waveform, contains all information about the speaker, and each step in the extraction process can only reduce the mutual information or leave it unchanged. The objective of the feature extraction is to reduce the dimension of the extracted vectors and thereby reduce the complexity of the system. The main task for the feature extraction process is to pack as much speaker-discriminating information as possible into as few features as possible.

The choice of features in any proposed ASV system is of primary concern, because if the feature set does not yield sufficient

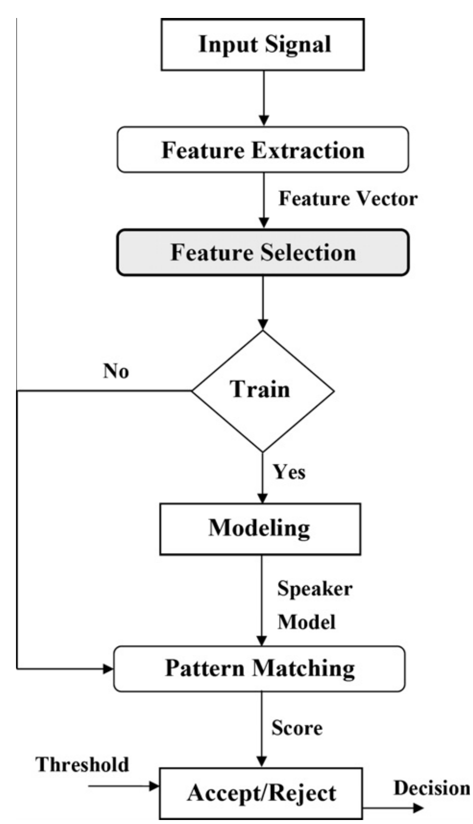

Fig. 1. Overview of the speaker verification process (Nemati et al., 2008).