Chapter 3.13
Differentially Fed Artificial Neural Networks for Speech Signal Prediction

Manjunath Ramachandra Iyer
Banglore University, India

ABSTRACT

Speaker authentication has become increasingly important. It goes with the other forms of security checks such as user login and personal identification number and has a say in the final decision about the authenticity. One of the issues with the authentication algorithms is that the automated devices that take the call have to work with a limited data set. In this chapter, a new class of intelligent element called differentially fed artificial neural network has been introduced to predict the data and use it effectively. It keeps the model simple and helps in taking online and crisp decisions with the available limited data.

INTRODUCTION

Online authentication of speakers has turned into a challenging task, especially with the kind of decisions to be taken based on the speaker authentication. This calls for an intelligent element that can make use of minimum data set and arrive at the meaningful conclusion.

Artificial neural networks can be conveniently used for pattern matching with a known set of data. The conventional neural networks, however, suffer with the drawback of a lengthy training period and reduced signal to noise ratio at the output. To address these issues, a differential feedback from the output of the network to the input is suggested and discussed. Intuitively, appropriate feedback from the output to the input of a system results in change in the system behavior, including the improved stability and noise immunity. A differential feedback ends up in more interesting properties. They require a reduced training set and provide better noise immunity. Some of the features are unique to these networks that are not found in a conventional neural network.

With the differentially fed neural network in place, a portion of the challenge data is taken for training data. The other part is synthesized

based on this. It is then matched with the actual data and the difference is computed. By setting threshold on difference, the speaker identity may be endorsed.

The knowledge base queries would consist of audio or speech samples. Sophisticated pattern matching algorithms would be required to generate the match online. Based on the outcome of the match, decisions would be taken. Such knowledge bases would be used in banks, industries, airport surveillance systems, crime detection and prevention, and so forth. It can prevent unauthorized entries. Online evaluation of the patterns is the biggest challenge. It calls for the usage of sophisticated algorithms. An intelligent algorithm based on differential feedback is discussed in this chapter.

BACKGROUND

There are various techniques for speaker identification and verification. The model-based algorithms with built-in auto regression can work with limited available data. Hence the stress in this chapter is for the data prediction models. In this section, different models are introduced for speaker verification. Basically two processes are involved with speech signal processing: speaker identification and verification.

Speaker identification is basically the task of determining who is speaking from a set of known voices or speakers. It is a pattern-matching problem. To perform identification, the system must perform a 1:N classification. Generally it is assumed the unknown voice comes from a known set of speakers. Often artificial neural network classifiers are used to accomplish this task. The matching of signal can happen with respect to a scale or shape of the waveform and not necessarily with respect to the numerical values.

Speaker verification or speaker authentication or detection is the task of determining whether a certain speech sample really comes or originates from a certain individual.

In another method of classification, there will be a closed set of users. It would be required to pick up one speaker out of a set. On the other hand, in an open set the speaker need not match with any of the known speakers.

Speaker identification involves the process of recognizing the speaker based on the information extracted from a part of the speech signal. The remaining part may be used to verify the authenticity. There are two classes of identification techniques: text dependent and text independent recognition. In a text dependent technique, the speaker is urged to pronounce a predetermined set of sentences. The text in both training and testing is the same or it is known in advance. In a text independent technique, the speaker identification happens without restricting the pronunciations of the speaker. The latter technique is more complex, as the target matching data will not be readily available. There will not be a known pattern of the spoken string. It has to be synthesized based on some other set of speech spoken by the speaker earlier. This chapter concentrates on the text dependent technique.

Speaker recognition system consists of two steps including feature extraction and feature matching. Feature extraction refers to parameterize the speech sample that is required later for feature matching. Feature matching involves the comparison of the features associated with a certain speech signal with the stored features and thereby identifies the speaker.

A speaker recognition system works on phases—the training phase and the deployment phase. During the training phase, the speakers, who are supposed to be recognized in future, are required to render voice samples. Based on these samples, individual speech models are built. It involves the extraction of the voice features and their storage. The deployment phase involves the classification of features of a certain speech signal into one of the sample sets. One of the major problems with the deployment phase is that the training data and the challenge data do not match most of the time. It is because the characteristics