Applications of Error Back-Propagation to Phonetic Classification

Authors: Hong Leung, Victor W. Zue

Abstract: This paper is concerced with the use of error back-propagation in phonetic classification. Our objective is to investigate the ba(cid:173) sic characteristics of back-propagation, and study how the frame(cid:173) work of multi-layer perceptrons can be exploited in phonetic recog(cid:173) nition. We explore issues such as integration of heterogeneous sources of information, conditioll~ that can affect performance of phonetic classification, internal representations, comparisons with traditional pattern classification techniques, comparisons of differ(cid:173) ent error metrics, and initialization of the network. Our investiga(cid:173) tion is performed within a set of experiments that attempts to rec(cid:173) ognize the 16 vowels in American English independent of speaker. Our results are comparable to human performance.

Early approaches in phonetic recognition fall into two major extremes: heuristic and algorithmic. Both approaches have their own merits and shortcomings. The heuristic approach has the intuitive appeal that it focuses on the linguistic informa(cid:173) tion in the speech signal and exploits acoustic-phonetic knowledge. HO'fever, the weak control strategy used for utilizing our knowledge has been grossly inadequate. At the other extreme, the algorithmic approach relies primarily on the powerful con(cid:173) trol strategy offered by well-formulated pattern recognition techniques. However, relatively little is known about how our speech knowledge accumulated over the past few decades can be incorporated into the well-formulated algorithms. We feel that artificial neural networks (ANN) have some characteristics that can potentially enable them to bridge the gap between these two extremes. On the one hand, our speech knowledge can provide guidance to the structure and design of the network. On the other hand, the self-organizing mechanism of ANN can provide a control strategy for utilizing our knowledge.

In this paper, we extend our earlier work on the use of artificial neural networks for phonetic recognition [2]. Specifically, we focus our investigation on the following sets of issues. First, we describe the use of the network to integrate heterogeneous sources of information. We will see how classification performance improves as moreError Back-Propagation to Phonetic Classification207information is available. Second, we discuss several important factors that can sub(cid:173) stantially affect the performance of phonetic classification. Third, we examine the internal representation of the network. Fourth, we compare the network with two traditional classification techniques: K-nearest neighbor and Gaussian classifica(cid:173) tion. Finally, we discuss our specific implementations of back-propagation that yield improved performance and more efficient learning time.