

Continuous speech recognition with sparse coding

W.J. Smit ^{a*}, E. Barnard ^b

a University of Pretoria, Pretoria 0001, South Africa

b Meraka Institute, P.O. Box 395, Pretoria 0001, South Africa

Received 30 July 2007; received in revised form 9 June 2008; accepted 15 June 2008

Available online 21 June 2008

Abstract

Sparse coding is an efficient way of coding information. In a sparse code most of the code elements are zero; very few are active. Sparse codes are intended to correspond to the spike trains with which biological neurons communicate. In this article, we show how sparse codes can be used to do continuous speech recognition. We use the TIDIGITS dataset to illustrate the process. First a waveform is transformed into a spectrogram, and a sparse code for the spectrogram is found by means of a linear generative model. The spike train is classified by making use of a spike train model and dynamic programming. It is computationally expensive to find a sparse code. We use an iterative subset selection algorithm with quadratic programming for this process. This algorithm finds a sparse code in reasonable time if the input is limited to a fairly coarse spectral resolution. At this resolution, our system achieves a word error rate of 19%, whereas a system based on Hidden Markov Models achieves a word error rate of 15% at the same resolution. © 2008 Elsevier Ltd. All rights reserved.

Keywords: Sparse coding; Spike train; Speech recognition; Linear generative model