Accepted Manuscript

Automatic context window composition for distant speech recognition

Mirco Ravanelli, Maurizio Omologo

PII: S0167-6393(18)30012-8
DOI: 10.1016/j.specom.2018.05.001
Reference: SPECOM 2563

To appear in: *Speech Communication*

Received date: 8 January 2018
Revised date: 14 March 2018
Accepted date: 8 May 2018

Please cite this article as: Mirco Ravanelli, Maurizio Omologo, Automatic context window composition for distant speech recognition, *Speech Communication* (2018), doi: 10.1016/j.specom.2018.05.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Automatic context window composition for distant speech recognition

Mirco Ravanelli, Maurizio Omologo
Fondazione Bruno Kessler, Trento, Italy

Abstract

Distant speech recognition is being revolutionized by deep learning, that has contributed to significantly outperform previous HMM-GMM systems. A key aspect behind the rapid rise and success of DNNs is their ability to better manage large time contexts. With this regard, asymmetric context windows that embed more past than future frames have been recently used with feed-forward neural networks. This context configuration turns out to be useful not only to address low-latency speech recognition, but also to boost the recognition performance under reverberant conditions.

This paper investigates on the mechanisms occurring inside DNNs, which lead to an effective application of asymmetric contexts. In particular, we propose a novel method for automatic context window composition based on a gradient analysis. The experiments, performed with different acoustic environments, features, DNN architectures, microphone settings, and recognition tasks show that our simple and efficient strategy leads to a less redundant frame configuration, which makes DNN training more effective in reverberant scenarios.

Keywords: Distant Speech Recognition, Deep Learning, Context Window, Reverberation

1. Introduction

Distant Speech Recognition (DSR) represents a fundamental technology towards flexible human-machine interfaces. There are indeed various real-life situations where DSR is more natural, convenient and attractive than traditional

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/6960500

Download Persian Version:

https://daneshyari.com/article/6960500

Daneshyari.com