World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:14, No:10, 2020

Reed: An Approach Towards Quickly Bootstrapping Multilingual Acoustic Models

Authors : Bipasha Sen, Aditya Agarwal

Abstract : Multilingual automatic speech recognition (ASR) system is a single entity capable of transcribing multiple languages sharing a common phone space. Performance of such a system is highly dependent on the compatibility of the languages. State of the art speech recognition systems are built using sequential architectures based on recurrent neural networks (RNN) limiting the computational parallelization in training. This poses a significant challenge in terms of time taken to bootstrap and validate the compatibility of multiple languages for building a robust multilingual system. Complex architectural choices based on self-attention networks are made to improve the parallelization thereby reducing the training time. In this work, we propose Reed, a simple system based on 1D convolutions which uses very short context to improve the training time. To improve the performance of our system, we use raw time-domain speech signals directly as input. This enables the convolutional layers to learn feature representations rather than relying on handcrafted features such as MFCC. We report improvement on training and inference times by atleast a factor of 4x and 7.4x respectively with comparable WERs against standard RNN based baseline systems on SpeechOcean's multilingual low resource dataset.

Keywords : convolutional neural networks, language compatibility, low resource languages, multilingual automatic speech recognition
Conference Title : ICSRLP 2020 : International Conference on Speech Recognition and Language Processing
Conference Location : Dubrovnik, Croatia
Conference Dates : October 01-02, 2020