

SAARLAND UNIVERSITY
UNIVERSITY OF LORRAINE

Abstract

MULTISPEECH Team

INRIA, LORIA, University of Lorraine (France)

M.Sc. in Language and Communication Technologies (LCT)

Leveraging Word Contexts in Wikipedia for Out-of-Vocabulary Proper Nouns Recovery in Speech Recognition

by Badr ABDULLAH

Automatic Speech Recognition (ASR) systems are usually trained on static data and a finite vocabulary. When a spoken utterance contains out-of-vocabulary (OOV) words, ASR systems misrecognize these words as in-vocabulary words with similar acoustic properties, but with entirely different meaning. The majority of OOV words are information-rich proper nouns (e.g., person names, geographic locations, commercial products) that are vital to spoken content understanding. Therefore, failing to recognise OOV words has a significant adverse impact on many downstream applications such as spoken document indexing and translation.

In this thesis, we approach this problem by dynamically extending the ASR vocabulary based on the context obtained from the ASR initial hypothesis. In other words, given the in-vocabulary transcription of a spoken utterance, the goal is to retrieve a ranked list of OOV proper nouns that are likely relevant to the spoken context, add words in this list to the ASR lexicon, and perform a second-pass decoding with the ASR system. To this end, we explore different techniques that leverage topical contexts of OOV words in Wikipedia to develop neural models for OOV words retrieval. Our experiments show that document-specific extension of the ASR vocabulary improves the performance of the speech recognizer.