Contextual Evidence and System Combination for Solving the Out-of-Vocabulary Word Problems in Speech Recognition

Long Qin

Committee:
Alex Rudnicky (chair)
Alan Black
Florian Metze
Mark Dredze (Johns Hopkins Univ.)

Abstract

Out-of-vocabulary (OOV) words are unknown words that appear in the testing speech but not in the recognition vocabulary. They are usually important content words, such as names and locations. But most speech recognition systems only recognize words in a fixed finite vocabulary. When encountering OOV words, such systems cannot identify OOV words, but mis-recognize them as in-vocabulary (IV) words. Furthermore, errors made on OOV words also affect the recognition accuracy of IV words. Speech recognition systems in which OOV words can be detected and recovered are therefore of great interest.

Current OOV solutions only collect evidence for individual words sometimes together with their surrounding words. This is much simpler than how human identifies OOV words. When we hear an OOV word, besides the uncertainty about individual OOV words, we often rely on broad contextual evidence especially syntactic and semantic evidence to verify whether what we hear is reasonable in a certain context. Furthermore, if other audiences also don't know that word, we are more confident that it is an OOV word. Therefore, we propose to improve the state of the art OOV word detection and recovery performance by applying broad contextual evidence and system combination. Specifically, we will investigate the use of various contextual evidence derived from the acoustic, lexical, language model, syntactic and semantic features. We will also explore various ways of utilizing multiple systems, such as combining multiple systems' features or outputs and using mixed types of sub-lexical units in one system. Like human, our system utilizes numerous clues not only collected from other words but also from other systems, thus it can perform much better than conventional methods. Finally, with the ability of identifying OOV words, speech recognition systems will be more robust and have broader applications in real life.

http://www.cs.cmu.edu/~lqin/cmu_files/proposal.pdf