Abstract

Keyword spotting (KWS) is a task for automatically detecting keywords of interest from continuous speech, which has been utilized in a broad variety of real life applications. A majority of such applications relate to audio indexing and speech data mining. For example, finding courses related to "speech recognition" from online lecture provider Coursera, where large amount of lectures are available in video format, instead of the traditional text format. Another portion of applications that is becoming increasingly important is the wake-up word, where keyword spotting is performed to activate a device or initiate a voice interaction interface. For example, Google's voice search features the keyword "Okay Google" where users can simply say "Okay Google" to initiate Google's voice search.

In this talk we propose to look at the low resource aspect of keyword spotting. We first look into the low language resource aspect of keyword spotting, where only limited language resource is available for training a keyword spotter. We show that a rich lexicon is critical for large vocabulary speech recognition (LVCSR)-based keyword spotting, and we propose techniques to improve upon a limited lexicon. We then look into the low computation resource aspect of keyword spotting, and we propose new keyword spotting algorithms that are suitable for computation constrained devices, such as mobile phones and tablets.