Joint CQSE and CASTS Seminar

Weekly Seminar
Apr. 15, 2016 (Friday)

TIME   Apr. 15, 2016, 14:30 ~ 15:30
TITLE   Spoken Content Retrieval - Beyond Cascading Speech
            Recognition with Text Retrieval
SPEAKER   Prof. Hung-yi Lee
            Department of Electrical Engineering,
            National Taiwan University
PLACE   Rm716, CCMS & New Physics Building, NTU

Abstract

Today the multimedia content over the Internet has become a part of human life, and the spoken part in multimedia content very often describes the core concept of the multimedia. Although the public search engines are very successful in searching for web pages based on the text, searching directly over the spoken content is still difficult. However, retrieval directly over the spoken content not only eliminates the needs to provide text information for the multimedia for indexing purposes, but can locate the exact utterance for the desired information.

Over the last decade, spoken content retrieval has achieved significant advances by primarily cascading speech recognition techniques with text information retrieval techniques. This approach works well when the recognition accuracy is high enough, but becomes less adequate for more challenging real-world tasks such as retrieving course lectures or telephone conversations with relatively low recognition accuracy, because the retrieval performance is highly dependent on the speech recognition accuracy.

Is it possible to have better retrieval results under relatively low speech recognition accuracy? In this talk, I am going to introduce some innovative approaches of spoken content retrieval beyond the mentioned cascading framework with retrieval performance shown to be less constrained by speech recognition accuracy. For example, the retrieval system can exploit the information in speech signals not present in the output of the standard speech recognition modules, speech recognition system can be redesigned to have better retrieval performance or even be eliminated from the spoken content retrieval process, the users can navigate across the desired spoken content with dialogues, and the search results can be visualized in an intuitive interface, etc..