Language Technologies Institute

Thesis Defense

GHC 4405| Friday, October 5, 2018| 11:30 am

Polyphonic Sound Event Detection with Weak Labeling

Yun Wang

COMMITTEE:

Florian Metze, (Chair)

Alex Hauptmann

Alex Waibel

Aren Jansen
(Google, Inc.)

Abstract

Sound event detection (SED) is the task of detecting the type and the onset and offset times of sound events in audio streams. It is useful multimedia retrieval, surveillance, etc. SED is difficult because sound events exhibit diverse temporal and spectral characteristics, and because they can overlap with each other.

Ideally, SED systems should be trained with strong labeling, which provides the type, onset time and offset time of each sound event occurrence. However, such labeling is formidably tedious to produce by hand. Current research on SED often uses weak labeling. This thesis deals with two types of weak labeling: presence/absence labeling, which only states which types of events are present in each recording without any temporal information, and sequential labeling, which only provides the order of sound events, but without timestamps. Even if the training data is weakly labeled, we still want our SED systems to localize the sound events in time.

SED with presence/absence labeling is usually treated as a multiple instance learning (MIL) problem, which requires a pooling function. In this thesis, we compare five pooling functions both theoretically and empirically, and establish the linearly weighted softmax pooling function as the optimal. Using this function, we build a state-of-the-art network that not only recognizes the types of sound events, but also localizes them temporally.

SED with sequential labeling has not received much attention. In this thesis, we propose a novel modified connectionist temporal classification (CTC) framework, which successfully makes use of the extra temporal information in sequential labeling compared with presence/absence labeling.

Transfer learning is a popular technique to deal with insufficient training data. In this thesis we extract features from two neural networks trained for out-of-domain tasks, and show that these features can improve the SED performance when the training corpus is small.

http://www.cs.cmu.edu/~yunwang/papers/cmu-thesis.pdf