Université de Lorraine - IDMC
Masters SC & TAL – M1
2018-2019

Fiche de projet tutoré / Project form

Transcribing out-of-Vocabulary words for Automatic Speech Recognition

Encadrement / Supervisors

1. équipe, laboratoire / team, lab : Equipes SMarT et Multispeech

1. encadrant·e principal·e (nom, email) / main supervisor (name, email) : Langlois David (david.langlois@loria.fr)

2. autres encadrant·es / other supervisors : Jouvet Denis (denis.jouvet@loria.fr)

Description / Description

1. projet global/global project

A speech recognition system aims at producing the sequence of words corresponding to an input speech signal.

Such system uses a close (limited) vocabulary of several hundreds thousand of words. Actually, when the acoustic signal contains out-of-vocabulary words (not in the initial vocabulary), the system does not detect that the speech segment corresponds to an unknown word, but it selects words from its initial vocabulary and outputs them. This leads to speech transcription errors. Such errors are dramatic because they create additionnal errors (Gauvain *et al.*, 2005) (Rosenfeld, 1995).

With the rise of the deep neural approach, speech recognition systems can now output a sequence of letters (Graves and Jaitly, 2014) with word separators. But this remains unsatisfying because a proposed sequence of letters may not match with true existing words.

Other approaches use a « loop » of phonemes (Messina and Kermorvant, 2014) to generate sequences of words and phonemes; a sequence of phonemes is considered as an unknown word. But this requires to generate an orthographic form for the sequence of phonemes, as can do a human being.

This project aims at studying research works which propose such a generation technique, and at experimenting/comparing several methods.

2. biblio. UE 705 (semestre 7)

The bibliography step will consist in studying scientific articles about this topic, starting with the initial article by Bisani and Ney (Bisani and Ney, 2005).

3. realisation. UE 805 (semestre 8)

This step will consist in creating a program which takes as input a sequence of phonemes,

and which proposes a corresponding sequence of words. Skills of the supervisors, and existing resources in SMarT and Multispeech teams will help students.

Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization

A computer will be be necessary to work. The students will work at Loria laboratory.

Livrables et échéancier / Deliverable and schedule

Delivrable 1 (semester 7, december) : state of the art about processing out-of-vocabulary words in automatic speech recognition.
The supervisors will help the students by proposing a strategy to find and select articles to be read.

Delivrable 2 (semester 8, march) : prototype for transcribing a sequence of phonemes into a sequence of letters. In this step, we will assume that the sequence of phonemes corresponds to one single unknown word.

Delivrable 3 (semester 8, may) : prototype for transcribing a sequence of phonemes into a sequence of letters (and word boundaries). In this step, we will assume that the sequence of phonemes correspond to one or several unknown or/and known words.

Bibliographie /References (max. 4-5)

Bisani, M., & Ney, H. (2005). Open vocabulary speech recognition with flat hybrid models. In *Ninth European Conference on Speech Communication and Technology*.
Gauvain, J. L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., Lamel, L., & Schwenk, H. (2005). Where are we in transcribing French broadcast news?. In *Ninth European conference on speech communication and technology*.
Graves, A., & Jaitly, N. (2014, January). Towards end-to-end speech recognition with recurrent neural networks. In *International Conference on Machine Learning* (pp. 1764-1772).
Messina, R., & Kermorvant, C. (2014, April). Over-generative finite state transducer n-gram for out-of-vocabulary word recognition. In *Document Analysis Systems (DAS)*, 2014 11th IAPR International Workshop on (pp. 212-216). IEEE.
Rosenfeld, R. (1995). Optimizing lexical and n-gram coverage via judicious use of linguistic data. In *Fourth European Conference on Speech Communication and Technology*.