



Institut de Recherche
en Informatique de Toulouse

Stage Master 2 / Ingénieur 3ème année
2021-2022

Discriminative sequence training in end-to-end automatic speech recognition

Keywords: Automatic Speech Recognition, Finite State Transducers

In the SAMoVA team, we develop Automatic Speech Recognition (ASR) models mostly for French [1]. Recently, we are using so-called end-to-end models, involving recurrent neural networks and transformers [2].

These models are usually trained with the standard supervised classification objective function, cross-entropy (CE). This loss function evaluates the model predictions one by one, and does not take into account the fact that the desired output is a sequence of symbols (words or characters). Some objective functions do, on the contrary, such as the Maximum Mutual Information (MMI) criterion. MMI involves the use of Finite State Transducers (FST), which are an efficient way to computationally model pronunciation lexicons and language models altogether. Sequence-based criteria have already been used in ASR [3], but not in a straight-forward seamless way in deep learning frameworks such as Pytorch so far.

The main objective of the internship is to train ASR models, with this kind of objective functions. Recently, MMI Pytorch implementations have been proposed, which will be a good starting point [4].

Qualifications

- Master 2 or Engineering school 3rd year in computer science
- Machine learning, data science background
- Proficiency in Python, and experience in Pytorch is a plus

Place: Toulouse, IRIT, Université Paul Sabatier

Duration: 5 to 6 months

Supervisor: Thomas Pellegrini (MdC UT3)

Contact: thomas.pellegrini@irit.fr, <https://www.irit.fr/~Thomas.Pellegrini/>

Please send me an email with your CV, briefly explaining why the topic is of interest to you.

Short bibliography

[3] H. Hadian, H. Sameti, D. Povey, et al. End-to-end Speech Recognition Using Lattice-free MMI. In Proc. Interspeech, Hyderabad, Sept. 2018

<http://www.interspeech2020.org/uploadfile/pdf/Thu-3-5-4.pdf>

[4] PyChain: A Fully Parallelized PyTorch Implementation of LF-MMI for End-to-End ASR, Yiwen Shao, Yiming Wang, Daniel Povey and Sanjeev Khudanpur. In Proc. Interspeech, Oct. 2020, Shanghai (online)

<http://www.interspeech2020.org/uploadfile/pdf/Mon-2-2-5.pdf>

Team publications

[1] Plateforme de Parole ATYpique: <https://paty.irit.fr/demo/>

[2] L. Gelin, M. Daniel, J. Piquier, T. Pellegrini, End-to-end acoustic modelling for phone recognition of young readers. In Speech Communication, Nov. 2021,

<https://arxiv.org/pdf/2103.02899>