LU TALKS

Language
Understanding
λ(μ) Laboratory

Improving Automatic Speech Recognition by Joint Training with Speaker Recognition and Translation Task

Chairperson

Dr. Ye Kyaw Thu
(Lab. Leader)
yktnlp@gmail.com

January 29
2022

11:00 am
(Myanmar Time)
13:30 pm
(Japan Time)

Our Guest Speaker

Kak Soky
Kyoto University, School of Informatics
Department of Intelligence Science and Technology,
Speech and Audio Processing Lab
E-mail: soky@sap.ist.i.kyoto-u.ac.jp

Biography

Kak Soky is a third-year Ph.D. student working with the Speech and Audio Processing Laboratory (Kawaharalab) at Kyoto University, Japan. His research interests lie in the application of machine learning methods for speech and language tasks, mainly in Automatic Speech Recognition (ASR). Recently he has been working on improving the ASR system by joint training with speaker recognition and translation task. He received a BSC from Svay Rieng University, Cambodia, 2010, MSC in IT Engineering from Royal University of Phnom Penh, Cambodia, in 2016. From 2012 to 2019, he was an adjunct lecturer at Svay Rieng University and National Institute of Posts, Telecoms and ICT (NIPTICT), Cambodia. From 2016-2017, he was an internship researcher at National Institute of Information and Communications Technology (NICT), Japan. Currently, he is an officer at the Ministry of Education Youth and Sports, Cambodia.

Abstract

Recently, automatic speech recognition (ASR) can reach up to a human level in high-resource languages, whereas, in low-resource languages, the performance is still not satisfying because of the lack of resources and other language processing tools. In this work, we create a spontaneous speech translation corpus of Khmer to English and French, which has three components: source speech, transcription, and translation text. Furthermore, we propose two joint training methods to enhance the Khmer ASR performance using a single Transformer-based model. First, joint training of speaker recognition (SRE) and ASR, which uses the speaker information of SRE output as a speaker embedding to integrate with the ASR decoder. Second, joint training of ASR and machine translation (MT) or speech translation (ST), which performs a simultaneous MT/ST of high-resource language and ASR of a low-resource language and then integrates them via cross-attention in a single ASR decoder.

Online Seminar
Via
Microsoft Team