



Combining Pretrained Language Model for Taiwan-accented Speech Recognition

Tenure-track Researcher

Dr. Sheng Li

BIOGRAPHY

Sheng Li received his B.S. and M.E. degrees in 2006 and 2009 from Nanjing University, Nanjing, China, and his Ph.D. from Kyoto University, Kyoto, Japan, in 2016. From 2009 to 2012, he worked at the Chinese Academy of Sciences researching speech technology-assisted language learning. From 2016 to 2017, he worked as a researcher at Kyoto University, studying speech recognition systems for humanoid robots. In 2017, he joined the National Institute of Information and Communications Technology, Kyoto, Japan, as a researcher working on speech recognition. He served as workshop co-organizer and session chair in interspeech2020, coling2022, odyssey2022, ACM Multimedia Asia2023. He is a member of the Acoustic Society of Japan (ASJ), the International Speech Communication Association (ISCA), and IEEE. He is now a member of the Speech, Language, and Audio (SLA) Technical Committee for APSIPA. His research is focusing on next-generation multilingual speech recognition/translation/synthesis and security-aware speech processing technology.

Chairperson


Dr. Ye Kyaw Thu


✉ yekyaw.thu@nectec.or.th

ABSTRACT

Taiwan-accented speech bears similarities to the Mandarin Min dialect, but with substantial differences in vocabulary, which significantly impacts spoken language recognition outcomes. This talk concentrates on integrating pre-trained language models (PLMs) with state-of-the-art self-supervised learning (SSL)-based speech recognition systems for Taiwan-accented speech recognition tasks. We propose a progressive error correction process in tandem with recognition to fully exploit the autoregressive nature of PLM models. Experimental results demonstrate that our method effectively addresses recognition errors stemming from misspelled vocabulary in accented speech. Our proposed progressive approach achieves roughly a 0.5% improvement compared to the conventional method. Furthermore, we demonstrate that fine-tuning PLMs solely with the text from the accented dataset can enhance recognition performance, despite the limitations of accented speech resources.

 meet.google.com/rqn-vipt-vfn

 9:00 - 10:00 (Thailand Time)

 November 18, 2023

