

数理・情報のフロンティア
2021 年度採択研究代表者

2022 年度
年次報告書

樋口 陽祐

早稲田大学 大学院基幹理工学研究科
大学院生

言語表現の階層構造に基づく End-to-End 音声認識の研究

研究成果の概要

本研究では、言語に関する事前知識を利用し、高精度な音声認識システムを構築する手法の開発に取り組んでいる。当年度は、事前学習済みの大規模汎用言語モデルから得られる言語知識を音声認識の過程に取り入れることで、音声認識の精度を向上させることを試みた。

音声認識において正確なテキストを生成するには、出力単語間の依存性を捉えることが重要となるが、これを音声情報のみから抽出するのは容易でない。例えば、ある発話音声に対して「あめ」という音を認識するだけでは不十分であり、その音が「雨」と「飴」のどちらを意図しているのかについても、文脈に応じて推定する必要がある。これに対し、大規模言語モデルである BERT から得られる汎用的な言語知識を音声処理の過程に明示的に組み込むことで、文脈情報を効果的に扱える音声認識手法を提案した。また、汎用言語モデルと音声認識で扱われるテキスト形式の違い（語彙の違い、句読点の有無や大文字・小文字の区別など）を解消するための手法を検討した。様々な言語や発話スタイル、学習データ量を用いた音声認識実験において提案手法を評価した結果、従来手法よりも高い認識精度が得られることを確認した。

当該成果は、自然言語処理の主要会議である Empirical Methods in Natural Language Processing (EMNLP2022)の Findings、および音声処理の主要会議である IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2023)に採択されている。

【代表的な原著論文情報】

- 1) “Hierarchical Conditional End-to-End ASR with CTC and Multi-Granular Subword Units”, Proc. ICASSP, pp. 7797-7801, May 2022.
- 2) “BERT Meets CTC: New Formulation of End-to-End Speech Recognition with Pre-trained Masked Language Model”, Proc. Findings of EMNLP, pp. 5486-5503, December 2022.
- 3) “BECTRA: Transducer-based End-to-End ASR with BERT-Enhanced Encoder”, Proc. ICASSP, June 2023. (accepted)
- 4) “InterMPL: Momentum Pseudo-Labeling with Intermediate CTC Loss”, Proc. ICASSP, June 2023. (accepted)
- 5) “CTC Alignments Improve Autoregressive Translation”, Proc. EACL, May 2023. (accepted)