

数理・情報のフロンティア
2021 年度採択研究者

2021 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(博士課程)

言語表現の階層構造に基づく End-to-End 音声認識の研究

§ 1. 研究成果の概要

本研究では、言語情報に関する事前知識を活用し、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。一般的に、深層学習技術を用いた音声認識の学習には、大量の教師データ(音声発話とその書き起こし文)を必要とする。これに対し、当年度では、「文字」や「単語」といった文の構成単位に着目し、出力文を段階的に推定する手法を検討することで、音声認識性能を向上することを試みた。

提案手法は、モデルの中間層に対して補助的な損失を適用し、音声認識の学習を行う。このとき、入力に近い層からは文字といった細かい単位、出力に近い層からは単語といった粗い単位の系列が予測されるようにモデルを学習する。出力文の単位を段階的に上げることで、入力音声から出力文への変換過程が効果的に学習されることを期待する。英語音声認識タスクにおいて単語誤り率を評価した結果、提案手法により既存モデルよりも高い性能が得られることを確認した。また、発話スタイルやデータ量に依らず、提案モデルが有効であることを示した。

本研究成果は音声処理分野の国際会議 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2022) に採択された。

【代表的な原著論文情報】

- 1) “End-to-End 音声認識のための粒度の異なるサブワード単位に基づく階層的な条件づけ”, 報処理学会研究報告 (SLP), vol.2021-SLP-139, no.19, pp.1-6, Dec. 2021.
- 2) “粒度の異なるサブワード単位に基づく階層的な条件付き End-to-End 音声認識”, 日本音響学会研究発表会講演論文集, pp.955-956, March 2022.
- 3) “Hierarchical conditional end-to-end ASR with CTC and multi-granular subword units,” Proc. 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2022), pp. 7797-7801, May 2022.