



大規模言語モデルの統合による音声認識システムの改善

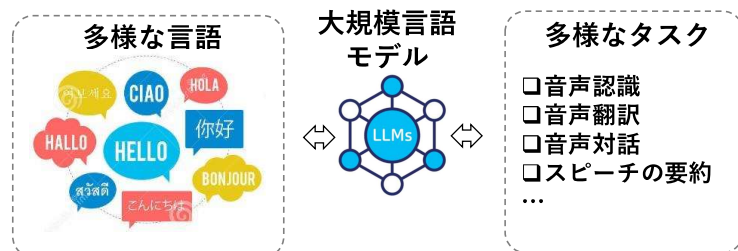
～ 人間レベルの精度を目指して～



概要

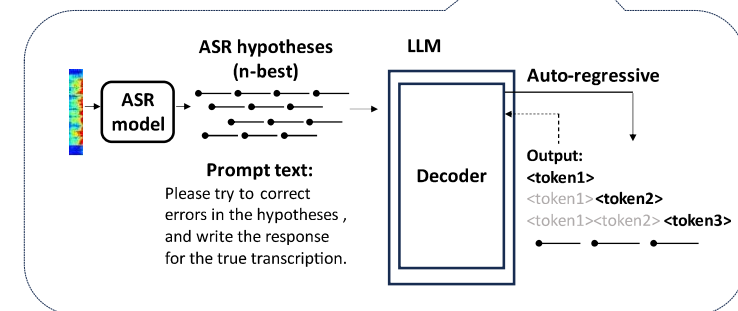
従来の自動音声認識 (ASR) システムは外部音響環境の影響を受け、認識精度が劣化する場合があります。この研究は、テキストを修正する最も強力な能力を持っている大規模言語モデル (LLM) を使用して、ASRシステムの認識結果を大幅に向上させることを目的 (人間レベルの精度) としています。

システム概要図



- LLMを利用してASR出力のエラーを修正します。提案手法では、LLMの高度な言語処理能力を活用し、ASRの結果が有効に改善されることが確認されました。

LLMに基づくエラー修正モデル



- また、文法やスペルの誤りの修正、その他のテキストの編集修正の多様なタスクにも使用できます。

特徴

- 大規模言語モデルを使用して従来の音声認識技術を強化
- 豊富な言語リソースに基づく言語モデルの活用により、人間レベルの精度を達成(文字エラー率1~2%)、リソースが少ない言語に対して、認識結果の精度が大幅に向上

ユースケース

- 音声認識、翻訳、対話への応用
- リソースの少ない言語の音声とビデオの記録をアーカイブおよび保護する

今後の展開

- 本手法の音声認識システムへの適用
- 音声通訳アプリや同時通訳システムへの展開

【お問合せ先】

ユニバーサルコミュニケーション研究所 先進的音声翻訳研究開発推進センター 先進的音声技術研究室 李 勝
Mail : sheng.li@nict.go.jp

NICTオープンハウス2024

Copyright © 2024 NICT All Rights Reserved.