**知能プログラミング演習I 第８回レポート（課題２）**

2024年 8月 1 日 学籍番号 34714037 氏名 加藤薫

調べた深層学習モデル：Whisper（音声認識，OpenAI）

１．モデルの原理

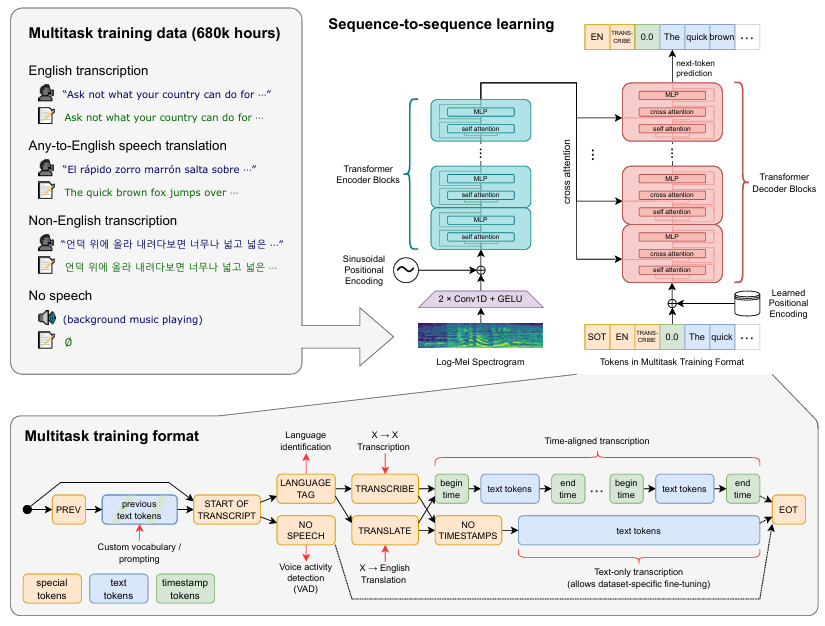


図 1**「Robust Speech Recognition via Large-Scale Weak Supervision」より引用**

WhisperはTransformerのエンコーダデコーダで構成される．入力音声は30秒のチャンクに分割され，ログメルスペクトログラムの形式で入力される．音声データをフーリエ変換して時間軸と周波数軸で表し，メル尺度に変換して，対数パワースペクトルに変換するとログメルスペクトログラムになる．エンコーダでログメルスペクトログラムから音声の特徴を抽出し，デコーダでエンコーダが抽出した特徴をもとに音声データをテキストに変換する．

２．利点および欠点，問題点

高精度かつ高速であるのでリアルタイムで文字起こしをすることができ，会議の議事録作成やインタビューの文字起こしに活用することができる．多言語対応しているため，異なる言語が飛び交う会議でも，すべての発言を正確に文字起こしすることができる．背景ノイズやアクセントに対しても高いロバスト性を持つ.

Whisperは幅広いデータセットでトレーニングされており，他のモデルよりも総合的に性能が良いが，特定のタスクにおいては他のモデルの方が優れている場合もある．英語に対する書き起こし精度は高いが，他言語に対する精度はデータセットの割合によって異なる．

利用する場合にはデータのセキュリティに注意する必要がある．

３．考察

自然言語処理の分野で登場したTransformerが音声認識でも精度が良くなるので，音声認識も自然言語処理と似たような処理であるか，Transformerが様々な分野で有用であると考えられる．Transformerで入力単語を埋め込み表現に変換した点と，音声認識でログメルスペクトログラムに変換した方が良いという点は似ている．深層学習を用いた音声認識は，カスタマーサポートの自動化や多言語会議の同時通訳など，利用できる場面も多いので，社会で大いに期待される技術であるといえる．

参考文献（Webページはすべて8月1日アクセス）

論文「Robust Speech Recognition via Large-Scale Weak Supervision」

Alec Radford Jong Wook Kim Tao Xu Greg Brockman Christine McLeavey Ilya Sutskever 著

[2212.04356 (arxiv.org)](https://arxiv.org/pdf/2212.04356)

ログメルスペクトログラムについて

[【音声認識】MFCCからlog-mel スペクトルで深層学習をしたら正解率が上がった #Python3 - Qiita](https://qiita.com/Oka_D/items/9afd11e501a1c540bc1b)

利点や欠点についてのWeb記事

[Whisperの高精度な理由とその活用法 - 議事録総合研究所 (securememo-cloud.com)](https://media.securememo-cloud.com/articles/why-whisper-is-highly-accurate-and-how-it-can-be-used/)

[Whisperモデルの解説とデモ (toolify.ai)](https://www.toolify.ai/ja/ai-news-jp/whisper%E8%A7%A3%E8%AA%AC-1063675#:~:text=Whisper%E3%81%AE%E5%86%85%E9%83%A8%E3%82%B7%E3%82%B9%E3%83%86%E3%83%A0%E3%82%A2%E3%83%BC%E3%82%AD%E3%83%86%E3%82%AF%E3%83%81%E3%83%A3%E3%81%AF%E3%80%81%E3%82%A8%E3%83%B3%E3%82%B3%E3%83%BC%E3%83%80%E3%83%BB%E3%83%87%E3%82%B3%E3%83%BC%E3%83%80%E3%83%99%E3%83%BC%E3%82%B9%E3%81%AE%E6%A7%8B%E9%80%A0%E3%82%92%E6%8E%A1%E7%94%A8%E3%81%97%E3%81%A6%E3%81%84%E3%81%BE%E3%81%99%E3%80%82,%E9%9F%B3%E5%A3%B0%E3%83%87%E3%83%BC%E3%82%BF%E3%81%AF%E3%83%AD%E3%82%B0%E3%83%A1%E3%83%AB%E3%82%B9%E3%83%9A%E3%82%AF%E3%83%88%E3%83%AD%E3%82%B0%E3%83%A9%E3%83%A0%E3%81%AE%E5%BD%A2%E5%BC%8F%E3%81%A7%E5%85%A5%E5%8A%9B%E3%81%95%E3%82%8C%E3%80%81%E3%82%A8%E3%83%B3%E3%82%B3%E3%83%BC%E3%83%80%E3%81%AB%E3%82%88%E3%81%A3%E3%81%A6%E9%9F%B3%E5%A3%B0%E3%81%A8%E3%83%86%E3%82%AD%E3%82%B9%E3%83%88%E3%81%AE%E5%AF%BE%E5%BF%9C%E9%96%A2%E4%BF%82%E3%81%8C%E3%83%9E%E3%83%83%E3%83%94%E3%83%B3%E3%82%B0%E3%81%95%E3%82%8C%E3%81%BE%E3%81%99%E3%80%82)

[音声認識AI「Whisper」とは？強み・活用方法・注意点を徹底解説！ - AI Market (ai-market.jp)](https://ai-market.jp/services/whisper/#Whisper-3)