

LLM を用いた海外講義動画の自動吹替/要約システム

佐藤 泰基[†] 吉田 雅裕[†]

[†]中央大学 国際情報学部 国際情報学科

1. はじめに

日本国内の学校教育や企業研修において、海外のMOOC (Massive Open Online Courses) などの講義動画を活用したオンデマンド教育が注目されている。海外講義動画の音声を日本語に吹替翻訳すると、動画音声の再生時間が変化し、動画と音声のミスマッチが発生することがある。そのため、海外講義動画の日本語化には人手による長時間の編集作業が必要となり、国内の教育現場への普及の妨げとなっている。そこで本稿では、大規模言語モデル (LLM) を活用した海外講義動画の自動吹替・要約システムを提案する。

2. 提案手法

本稿では、動画音声の再生速度の調整と、音声内容の要約を組み合わせて、元の講義動画から自動生成した日本語の音声を、動画内の適切なタイミングで再生する手法を提案する。まず、音声認識モデルを用いて動画音声の文字起こしを行い、LLM で吹替翻訳と要約を行う。次に、音声合成モデルを用いて日本語の音声を作成する。提案手法では、自動生成された日本語音声の再生時間に合わせて、講義動画の再生速度とタイミングを自動調整することで、動画と音声がずれない日本語の講義動画を実現する。

3. 実装

Python を用いて提案手法の実装を行った。音声認識モデルには Whisper, LLM には gpt-3.5-turbo-1106, 音声合成モデルには tts-1[2], Web フレームワークには Django を用いた。LLM の出力結果に基づいて音声を合成すると、チャプターごとの映像の再生速度が計算され、講義動画の再生時の映像制御に利用される。チャプターはトピックに基づいて LLM により分割されたもので、コンテキスト長が限られている LLM による翻訳や要約の品質を担保するために必要となる。

4. 評価

(1) チャプター分割の精度評価

LLM によるチャプターの分割が不適切であると、吹替および要約の品質に悪影響を及ぼす恐れがある。そこで、動画内のスライドの遷移タイミングをチャプター分割の正解としてチャプター分割の精度を評価した。神経科学を扱った講義動画 10 本[3]を対象に実験を行った結果、表 1 に示す結果が得られた。

表.1 チャプター分割の精度評価結果

	正解率	適合率	再現率	F1 値
結果	0.8169	0.5964	0.8717	0.7082

81%の正解率でチャプターの分割を行えることから、LLM によるチャプターの分割は適切であると言える。また、スライド遷移のタイミングを逃さないように再現率を高めているため、適合率は低下していることが分かる。

(2) 提案手法が生成した講義動画の再生速度
提案手法が生成した講義動画の映像の平均再生速度を算出したところ、表 2 に示す結果が得られた。

表.2 処理後の平均再生速度

	吹替	要約
平均再生速度	0.9376	3.5765

通常の吹替では元の動画と概ね同じ程度の再生速度であり、要約の場合は 3 倍程度の再生速度であった。

(3) 利用体験の定性及び定量評価

提案手法により自動生成された講義動画を視聴した被験者 13 名に講義動画の理解度テストとアンケート調査を行い、被験者から良好な結果が得られた。

5. おわりに

本稿では、海外講義動画を自動的に吹替翻訳し日本語の音声を挿入するシステムの提案および評価を行った。評価実験の結果、提案手法はスライドを使った講義動画に対して、適切なタイミングで日本語の音声を自動的に挿入できることを確認した。一方、提案手法は映像の再生速度を変更するため、映像表現に重きを置く講義動画には不向きであるため、今後の課題として解決していく予定である。

参考文献

- [1] M. Otani et al. “Video Summarization Overview,” arXiv:2210.11707, 2022
- [2] OpenAI “Models,” <https://platform.openai.com/docs/models>, 参照 2023 年 12 月 22 日.
- [3] Neuroscientifically Challenged “2-Minute Neuroscience,” <https://neuroscientificallychallenged.com/videos>, 参照 2023 年 12 月 22 日.