

# LLM による対話を用いた講義動画検索

原田 優輝\*, 平川 翼, 山下 隆義, 藤吉 弘亘(中部大学), 遠藤 太一郎(カナメプロジェクト)

Preparation of Papers for Tokai-Section Joint Conference on Electrical, Electronics, Information and Related Engineering

Yuki Harada, Tsubasa Hirakawa, Takayoshi Yamashita, Hironobu Fujiyoshi (Chubu University), Taichiro Endo (Kaname Project)

## 1. 序論

本研究では、大規模言語モデルなどの学習済み深層学習モデルを使用して、動画授業を受講者の復習に役立てる方法を提案する。受講者は教師に直接質問することなく、このシステムを利用して簡単に疑問を解決することが可能である。また、文章での回答に加えて、該当する講義とそのタイムスタンプをシステムが推薦することで、受講者は膨大な時間の講義動画から疑問を解決するための箇所を見つけ出し、効率的に復習することができる。

## 2. 提案手法

本研究では、スライドを使用した音声解説形式の講義において長時間の講義でも受講者が効率的に復習できるようにすることを目的とする。具体的には、素朴な画像処理と大規模言語モデル (LLM) をはじめとする複数の深層学習モデルを組み合わせた、対話型の講義動画検索手法を提案する。

Fig. 1 に提案手法の概要を示す。まず、システム側において、講義動画に対する前処理として、口頭での解説内容の文字起こしおよび解説文の埋め込みベクトルのデータベース構築を行う。次に、受講者は講義内容に関する質問を対話形式の文章として入力する。入力された質問文は、前処理と同様のテキスト埋め込みモデルを用いてベクトル化する。質問文の埋め込みベクトルとデータベース内の解説文の埋め込みベクトルの類似度を計算することで類似する解説文を抽出する。この類似する解説文を LLM を用いた Retrieval-Augmented Generation (RAG) (1) に使用することで、自動的に質問に対する回答を生成する。さらに、生成された回答文の埋め込みベクトルを用いてデータベース内の類似文章検索を再度行い、類似する解説文に紐づけられた動画とそのタイムスタンプを受講者に提示する。

本手法により、受講者は膨大な時間の講義動画から疑問を解決するための箇所を見つけ出し、効率的な復習が可能となる。以下では、本手法内の各処理について詳細に述べる。

### <2・1>前処理によるデータベース作成

本節では、講義動画に対する前処理について述べる。まず、講義動画内のスライドの切り替わりを画像微分で検出し、スライドごとに講義内容をセグメント化する。次に、Fast Whisper (2) を使用して講義動画の講義動画の文字起こしを行い、各スライドの先頭にタイムスタンプを付与する。次に、テキスト埋め込みモデルである multilingual-e5-large (3) を用いて、文字起こしされたテキストを埋め込みベクトルへ変換する。このようにして生成されたタイムスタンプ付きテキストと埋め込

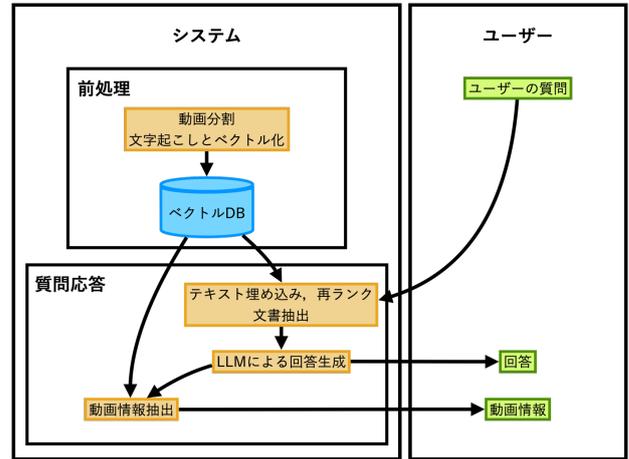


Fig.1. Overview of lecture video search by LLM

みベクトルをデータベースに保存する。

本データベースを事前構築することで、後続の RAG やベクトル検索において情報を検索し、関連する講義部分を特定することが可能となる。

### <2・2>LLM を用いた RAG による回答文生成

本節では、受講者から与えられた質問文から講義内容に沿った回答文を生成する処理手順について述べる。

まず、ユーザーから与えられた質問文を前処理と同様のテキスト埋め込みモデルである multilingual-e5-large でベクトル化する。このベクトルを基にデータベース内の埋め込みベクトルとの類似度計算を行い、最も類似したタイムスタンプ付きテキストを 10 件抽出する。次に、抽出された文書をクロスエンコーダモデルである japanese-reranker-cross-encoder-large-v1 (4) で再ランクする。クロスエンコーダでは、抽出された文書を与えられた質問と同時に入力し、より精度の高い類似度評価を行う。これにより、最も関連性の高い 3 件の文書が選定される。最後に、再ランクされた 3 件の文書をプロンプトに挿入し、LLM である LLaMA 3-70B (5) を使用して RAG を実行し、回答文を生成する。

### <2・3>生成された回答と動画の該当箇所の紐付け

本手法では、効率的な講義動画の復習のために、生成された回答文に関連する講義動画の箇所を受講者に提示する。本節では、回答文に該当する動画箇所の紐付け方法を説明する。

まず、生成された回答をテキスト埋め込みモデルでベクトル化する。このベクトルを基に、データベース内のタイムス

	強く 思う	そう 思う	普通	合計
学部生	21	11	2	34
社会人	1	2	1	4
合計	22	13	3	38

Fig.2. Percentage of Responses to Each Question

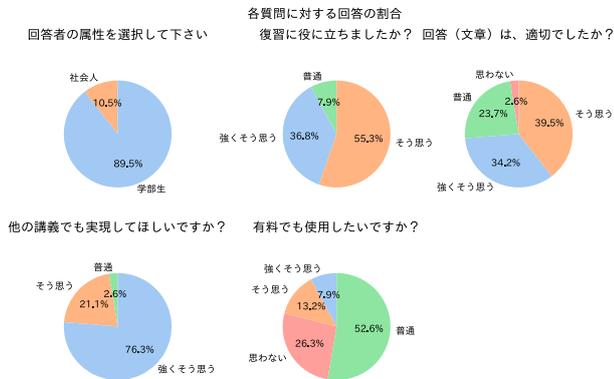


Fig.3. Percentage of Responses to Each Question

タンプ付きテキストから類似する文書を 10 件抽出する。その後、抽出された文書をクロスエンコーダで再評価する。再ランクされた文書の中から最も適切な文書を選定し、選定された文書に紐づけられた動画とそのタイムスタンプを事前の RAG で生成された回答文と共に受講者に提供する。

### 3. 提案システムの有効性評価

本手法の有効性を確認するため、実際の講義に本手法を用いたシステムを開発し、使用者に対してアンケート調査を実施する。受講者の内訳は学部生が 34 人、社会人が 4 人であり、受講者それぞれが希望する講義で調査を行った。

#### <3・1>提示する動画フレームの妥当性評価

まず、本システムを実際の講義復習ツールとして導入し、受講者に対して生成された回答に関連する動画フレームが適切であったかを実験した。Fig.2 に提示した動画フレームの妥当性に関するアンケートの結果を示す。この結果によると、38 人中 22 人の回答者が「強く思う」と回答し、35 人の回答者が「そう思う」以上の評価をしている。この結果から、動画への接続性は高く、本システムの動画検索機能の精度が十分であることが示された。

#### <3・2>復習の効果に関する評価

受講者に対するアンケート調査を通じて、本システムが講義の復習に効果的であったかを評価した。調査には先のアンケートと同様の 38 名の受講者が回答し、以下の 4 つの項目を評価した。

#### <項目>

1. 本システムは、復習に役に立ちましたか？
2. 本システムの回答（文章）は、適切でしたか？
3. 本システムを他の講義でも実現してほしいですか？
4. 本システムを有料でも使用したいですか？

質問については、以下の 4 段階で評価してもらった。

- (ア) 強く思う
- (イ) そう思う
- (ウ) 普通

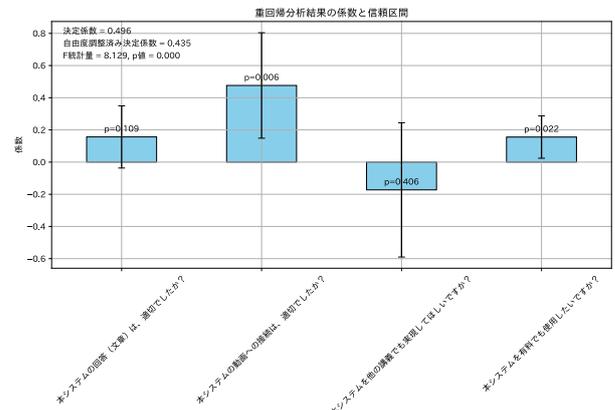


Fig.4. Coefficients and Confidence Intervals of Multiple Regression Analysis Results

#### (エ) 思わない

項目 1 では、本システムが講義の復習に役立つかどうかを評価した。Fig.3 に示すアンケートの結果、92.1%の回答者が本システムを復習に役立つと評価しており、本システムが授業の復習に有効であることが示された。さらに、復習の効果に対する要因を分析するために回帰分析を実施し、その結果を Fig.4 に示す。アンケート結果を 4 段階の量的データとして捉え、項目 1 を目的変数、それ以外の項目を説明変数として予測モデルを作成した。分析の結果、動画の接続性が復習の効果に正の影響を与えていることが示された（回帰係数: 0.4764, p 値: 0.006）。これにより、今後は動画へのアクセスの容易さの改善が求められることが示唆される。また、有料での使用意向も重要な因子であることが示されており（回帰係数: 0.1555, p 値: 0.022）、システムの付加価値の向上が望まれる。一方で、回答の適切さに関しては相関が見られるものの、統計的に有意ではないことが示された（p 値: 0.109）。

### 4. まとめ

本研究では、素朴な画像処理と大規模言語モデル (LLM) をはじめとする複数の深層学習モデルを組み合わせた、対話型の講義動画検索手法を提案した。アンケートによる有効性評価の結果、本システムが授業の復習に効果的であることを示した。また、重回帰分析の結果、動画の接続性が復習の効果に正の影響を与えることが示された。これにより、今後は動画へのアクセスの容易さの改善が求められることが示唆された。一方、回答の適切さに関しては統計的に有意な結果は得られなかった。総じて、動画への接続性が復習に効果的であることが明らかとなり、テキスト埋め込みや再ランクモデルなど大規模言語モデル以外の学習済みモデルも併用することが効果的であると考えられる。

#### 文献

- (1) P. Lewis, et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks, NeurIPS, 2020.
- (2) Systran: Faster Whisper transcription with CTranslate2, <https://github.com/SYSTRAN/faster-whisper>, 2024.
- (3) L. Wang, et al.: Multilingual E5 Text Embeddings: A Technical Report, arXiv preprint, 2024.
- (4) Y. Tateno: 日本語 Reranker 作成のテクニカルレポート, <https://secon.dev/entry/2024/04/02/080000-japanese-reranker-tech-report/>,

2024.

(5) Meta LLaMA Team ; Introducing Meta Llama 3: The most capable openly available LLM to date. <https://ai.meta.com/blog/meta-llama-3/>, 2024.