テキストデータの解析に基づく動画の生成

1 はじめに

近年, 動画配信サービスが盛んになり, 動画コンテンツの重要性が増している. その中には公知のミームを用いることで必要な素材の数を減らし, 容易に制作が可能なコンテンツが存在する.

本研究では, テキストから必要な情報を抽出し, 公知のミームを用いた動画を生成する手法を提案する. テキストからの動画の生成が容易になれば, 動画の生成に関して時間の節約や作成のコスト削減, 大量のコンテンツの作成が可能になる. 具体的には, テキストを解析し, それを元にした背景画像とミーム素材等を決定, これらをもとに動画の生成をする.

また, 様々な動画生成 AI が発表されている. 本研究で扱った猫ミームは, 動画素材が有限であるため, ストーリーを入力した際に素材の中から適切なものを選択することで動画の生成が容易になる. 加えて, 公知のミームを使用するため, 視聴者はそのミームに対して持つイメージをもとにストーリーを理解することができる.

2 要素技術

2.1 LLM

LLM (Large Language Model) とは, 大規模なデータセットを使用して訓練された自然言語処理のモデルを指す. 適切なデータを用いてファインチューニングすることで, 自然言語理解, 感情分析, 文章生成などのさまざまなタスクを実現できる.

2.2 GPT

GPT (Generative Pre-trained Transformer)[1] は, OpenAI が開発した言語モデルである. 自然言語や生成タスクにおいて, 高速で効率的な処理を実現し, 複雑な文脈を理解する能力を有する.

本研究では, ミームを決定する際や比較実験において, gpt-4o-mini というモデルを使用した.

2.3 Faiss

Faiss (Facebook AI Similarity Search)[2] は, 高次元ベクトルの効率的な類似性検索とクラスタリングのためのライブラリである. 主に大規模データセットに対する近似最近傍探索を目的としており, メモリ内での検索やインデックスの圧縮技術を活用して, 高速かつスケーラブルな処理を実現する. 本研究では, 背景画像の選択に使用した.

2.4 Gemini

Gemini[3] は, Google が開発した大規模言語モデルで, テキストや画像などの複数のデータ形式を処理できる. 文脈理解や応答精度に優れ, さまざまなタスクに対応できる高性能な AI システムである.

本研究では, テキストから情報を抽出する際に, 言語モデルとして gemini-1.5-flash を使用した.

2.5 SQL

SQL (Structured Query Language) は, リレーショナルデータベースのデータを操作するための言語である. データの検索, 追加, 更新, 削除などを行うための標準的な構文を提供する. 本研究では, MySQL を使用した.

2.6 Adobe After Effects (AE)

Adobe After Effects は, Adobe が提供する高度なデジタル映像編集・合成ソフトウェアである. 主にモーショングラフィックス, ビジュアルエフェクトを動画に加えることに利用され, 動画, テレビ番組, ウェブコンテンツ, 広告などで活用されている. After Effects は, 他の Adobe 製品 (Premiere Pro, Photoshop, Illustrator 等) との連携が強力である.

また, After Effects には, 多数の内蔵エフェクトがあり, これらを組み合わせることで様々な表現が可能である. さらに, プラグインやスクリプトの追加で機能を拡張し, 特定のニーズに合わせたカスタマイズが可能である.

2.7 ExtendScript

ExtendScript は, Adobe 製品向けの JavaScript ベースのスクリプト言語で, Adobe Effects, Photoshop, Illustrator などの自動化やカスタマイズを可能にする.

通常, ExtendScript は Visual Studio Code の拡張機能である ExtendScript. Debugger を使用する. これにより, プログラムでソフトの操作を制御し, 特定の作業フローに合わせた柔軟な解決策を構築できる.

2.8 ミーム

ミーム [4] とは, インターネット上で Web サイトや SNS を通して拡散され, 話題となった文章や画像, 動画のことである. 多くの人が共通で認識するネタやコラージュ画像などが含まれ, 文化的に人から人へと広がる現象を指す. 元々「ミーム」は, 進化生物学者リチャード・ドーキンスが命名した, 人から人へ広がる行動やアイデアの概念である.

3 提案手法

本研究では, テキストを解析し, 動画を自動生成する手法を提案する.

1. テキストから情報を抽出
2. 背景画像[1], 猫ミーム素材[2]に関するデータを入力した MySQL のデータベースを使用し, 適切なファイルパスを入手
3. ファイルパス, テキスト情報をもとに, ExtendScript を作成, 実行

テキストからの情報の抽出には gemini-1.5-flash を, 背景画像のファイルパスの決定には Faiss を, 猫ミーム素材のファイルパスの決定には gpt-4o-mini を使用した. gpt-4o-mini のパラメータは temperature を 0.5 に, gemini-1.5-flash は temperature を 1.0 に設定した.

3.1 テキストからの情報の抽出

以下に, テキストから情報を抽出する際に用いたプロンプトを示す. これは, gemini-1.5-flash というモデルを使用している.

プロンプト

日本語で書かれたこの文から, 時間, 場所, 登場人物の状態, 状況を簡単に説明する 10 文字程度のテキストの 4 つの情報を以下の形式で抽出してください. 時間, 場所, 登場人物の状態に関して仮にわからない場合, 不明と出力してください. テキストに関しては必ず出力してください.: {text}
時間:
場所:
登場人物の状態:
テキスト情報:

このプロンプトを使用して, 入力文から時間, 場所, 登場人物の状態, テキスト情報を入手する.

3.2 ファイルパスの決定

抽出した情報をもとに動画に使用する素材を決定する.

表 1, 2, 3 に背景画像に関する情報をまとめたデータベースの一部を示す. 表 images には, 画像 ID, 場所 ID, 時間 ID, ファイルパスが存在し, 1 つの場所に複数の時間が存在する形で保存している. これにより, 場所を決定した後に時間を決定するという形式で実装が可能になる.

表 4, 5, 6 にミーム素材に関する情報をまとめたデータベースの一部を示す. 表 Meme_Features にはミーム ID, 特徴 ID, ファイルパスが存在している. 同一ミーム ID が複数の特徴を持つ場合があり, また複数のミーム ID が同一の特徴を持つ場合もある.

images			
image_id	location_id	time_condition_id	file_path
1	1	3	/Users/...
2	1	2	/Users/...
3	1	1	/Users/...
4	2	3	/Users/...

表 1: 表 images の構成

locations	
location_id	location_name
1	ATMコーナー
2	アーケード商店街
3	アイランドキッチン
4	アジト

表 2: 表 locations の構成

[1] https://min-chi.material.jp

[2] https://sasalabo.net/2024/02/27/猫ミームとは/#google_vignette

time_conditions	
time_condition_id	time_and_condition
1	日中
2	夜
3	夕方
4	夜・照明OFF

表 3: 表 time_conditions の構成

Meme_Features		
meme_id	feature_id	file_path
1	1	/Users/...
2	2	/Users/...
2	3	/Users/...
3	3	/Users/...

表 4: 表 Meme_Features の構成

以下に背景画像のファイルパスの決定の方法を示す.
1. 抽出された場所名を入力し, ベクトル検索により適切な場所を出力
2. その場所に対応する ID を取得し, 抽出された時間を入力して, ベクトル検索により適切な時間を出力

以下にミーム素材のファイルパスの決定の方法を示す.
1. 抽出された特徴を入力し, GPT を使用して特徴の表から適切な状態の ID と名前を取得
2. その状態 ID を持つミーム ID を取得
3. 再度特徴を入力し, GPT を使用して, ミーム名を含めて再度適切なミームの ID と名前を取得

3.3 ExtendScript の作成, 実行

3.1 で得たテキスト情報と 3.2 で得たファイルパスを使用して動画を作成するためのスクリプトを生成する. 動画の長さ, 素材の位置やサイズ, テキストの表示時間等を事前に指定したテンプレートに対して, テキスト情報とファイルパスを入力することで, 動画を生成するための ExtendScript ファイルを出力させる. このコードを実行することで動画が生成される.

4 実験

本研究では, 2 つの実験を実施した.

1 つ目は, テキストから情報を抽出する際に, 使用する LLM を gpt-4o-mini と gemini-1.5-flash で比較する実験である. 同一のプロンプトを使用し, 同じ入力文から抽出された情報を 5 回試行で比較した.

Memes	
meme_id	meme_name
1	DJ猫
2	EDM猫
3	Girlfriend猫
4	olia猫

表 5: 表 Memes の構成

Features	
feature_id	feature_name
37	1匹の猫を叩く
2	EDM
9	いびきをかく
5	キーボード

表 6: 表 Features の構成

2 つ目は, 背景画像のファイルパスを決定する際に, gpt-4o-mini を使用した場合と Faiss を使用した場合で比較する実験である. gpt-4o-mini を使用する場合も Faiss と同様に場所を決定した後に時間を決定するように実装した. 異なる 5 つの場所と時間を入力し, 得られた場所名, 時間名を比較した. 本研究では SQL を使用しており, 画像 ID が決定すれば, それに対する場所名, 時間名, ファイルパスも一意に定まる.

4.1 実験 1: GPT

入力文: 勉強している時は、一休みすると気分が楽になる。
表 7 に 5 回試行の結果を示す.

表 7: 実験 1: GPT 5 回試行

時間	場所	登場人物の状態	テキスト情報
勉強中	不明	一休み中	気分が楽になる
勉強中	不明	一休み中	気分が楽になる
勉強中	不明	一休み中	気分が楽になる
勉強中	不明	一休み中	気分が楽になる
勉強中	不明	一休み中	気分が楽になる

4.2 実験 1: Gemini

入力文: 勉強している時は、一休みすると気分が楽になる。
表 8 に 5 回試行の結果を示す.

表 8: 実験 1: Gemini 5 回試行

時間	場所	登場人物の状態	テキスト情報
勉強中	不明	疲れている	気分転換が必要
勉強中	不明	疲れている	気分転換必要
勉強中	不明	疲れている	気分転換が必要
勉強中	不明	疲れている	気分転換必要
勉強中	不明	疲れている	一休みで気分転換

4.3 実験 2: GPT

表 9 に 5 つの入力を与えた結果を示す.

表 9: 実験 2: GPT

入力場所	入力時間	出力場所	出力時間
ファミレス	昼	レストラン	日中
図書館	夜	市立図書館	照明 ON
学校	昼休み	学校のベンチ	日中
街中	夕方	街中のビル	夕方
自宅	夜	家	夜・照明 OFF

4.4 実験 2: Faiss

表 10 に 5 つの入力を与えた結果を示す.

表 10: 実験 2: Faiss

入力場所	入力時間	出力場所	出力時間
ファミレス	昼	アジト	照明 ON
図書館	夜	図書室	夕方
学校	昼休み	学校のベンチ	日中
街中	夕方	都会の街中	夕方
自宅	夜	家	夕方

5 考察

実験 1 で, GPT を使用した場合と Gemini を使用した場合のいずれも, 時間と場所は同様の結果を示した. しかし, 登場人物の状態とテキスト情報に関しては, 両者で明確な違いが見られる. GPT を使用した場合は, 入力文にある情報をもとに, 登場人物の状態を「一休み中」, テキスト情報を「気分が楽になる」と出力している. 一方で Gemini の場合は, 勉強して一休みするという状況から登場人物が疲れていると想定したのか, 登場人物の状態を「疲れている」とした上で, テキスト情報で「気分転換が必要」という説明を入れている. Gemini の方が入力文をより理解していると考えられる.

実験 2 では, 「ファミレス」のような直接データベースに存在していないものを使用した. GPT はこれを正確に解釈し, 他の場所や時間に対しても正確に出力している. LLM を使用しているため, 高度な検索が可能になっていることがわかる. 一方で, Faiss を使用したベクトル検索に基づくファイルパスの決定は, 「ファミレス」などの存在していない単語や略称を処理できていない.

6 まとめと今後の課題

本研究では, テキストの解析を実施し, SQL を使用して背景画像と動画素材の決定をし, これらを使用した動画の生成という流れを自動化するスクリプトを作成し, 実行した.

今後の課題として, 背景画像の自動生成, ミーム素材の生成の検討が挙げられる.

本実験では, 元々インターネット上に存在する背景画像を使用したが, 文章から抽出した時間や場所の情報をもとに, 背景画像の自動生成をしたいと考えている. また, 本研究では 46 種類のミーム素材に特徴をラベル付けしているが, 十分なラベル付けができていなかったり, 登場人物の状態に合うミーム素材がなかったりすることがあったため, ミーム素材を生成することで解決できないかを検討していきたい.

参考文献

[1] OpenAI. Models. https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4. Accessed: 2024-12-1.

[2] masuidrive. 最近話題の vector search を実現する faissって何? #1. https://note.com/masuidrive/n/n5dc6da6dd2b6. Accessed: 2024-12-4.

[3] Google. Gemini api ドキュメント. https://ai.google.dev/gemini-api/docs?hl=ja. Accessed: 2024-12-04.

[4] ソフトバンクニュース. 【インターネットミーム】1分でわかるキーワード. https://www.softbank.jp/sbnews/entry/20230420_02, 2023.