

# 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、インターネット上にはさまざまなテキストデータが存在する。この中の文章から動画の生成が実現できれば、動画の生成に関して時間の節約、コスト削減に加えて大量のコンテンツの作成が可能になる。

本研究では、それらのテキストのうちストーリー性のある文章を取り上げ、テキストから必要な情報を抽出し、そこから動画を生成することを目的とする。文の解析を実施し、それをもとにした背景画像と動画素材の決定、これらの素材を使用した動画の生成という手順で進めていく。なお、動画生成するスクリプトを作成する上で簡略化のために、日常の出来事を背景画像、動画素材、テキスト情報等の単純な要素で構成する猫ミーム動画を扱った。

## 2 要素技術

### 2.1 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。主に映像の合成、モーショングラフィックス、ビジュアルエフェクトの制作に利用され、映画、テレビ番組、ウェブコンテンツ、広告など多岐にわたるメディアで活用されている。After Effects は、他の Adobe 製品（例えば、Premiere Pro, Photoshop, Illustrator）との連携が強力で、統合的なワークフローを提供する。

このソフトウェアは、タイムラインベースのインターフェースを持ち、ユーザーがエフェクトを適用したり、アニメーションを作成したりするのを容易にする。特に、キーフレームを使ったアニメーションの制御が強力で、精緻な動きの調整が可能である。また、3D 空間での作業やカメラの操作もサポートしており、より複雑なビジュアルエフェクトを作成できる。

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無限の表現が可能である。さらに、さまざまなプラグインやスクリプトを追加することで、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。

### 2.2 ExtendScript

ExtendScript は、Adobe 製品向けのスクリプト言語で、主に JavaScript ベースの記述方法を採用している。Adobe After Effects, Photoshop, Illustrator などのソフトウェアで利用され、これらのアプリケーションの自動化やカスタマイズを可能にする。ユーザーは ExtendScript を用いて、反復作業の自動化、複雑なタスクの簡略化、独自のツールやパネルの作成が容易に実行できる。

ExtendScript は、Visual Studio Code の拡張機能である ExtendScript Debugger で利用することが一般的である。これを利用することで、ユーザーはソフトの操作をプログラムで制御し、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。ExtendScript は、特にプロのクリエイターや開発者にとって、効率的なワークフローを実現するための強力な道具である。

### 2.3 GPT-3.5-turbo

GPT-3.5-turbo は、OpenAI が開発した言語モデルで、GPT-3.5 の改良版である。このモデルは、自然言語理解や生成タスクにおいて、高速で効率的な処理を実現する。GPT-3.5-turbo は、多くの情報を処理し、複雑な文脈を理解する能力を有する。

このモデルは、大規模なトレーニングデータと豊富な言語知識を基にしており、一般的な会話から専門的なトピックまで幅広い領域で対応が可能である。また、文法の正確さやコンテキストに即した回答の提供にも優れている。

## 3 提案手法

本研究では、文の解析を実施し、動画を自動生成する手法を提案する。

1. 文から情報を抽出
2. 背景素材、猫ミーム素材に関する情報をまとめた csv ファイルを利用し、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、テンプレートに当てはめて ExtendScript を作成、実行

文からの情報の抽出とファイルパスの決定の 2 つに GPT-3.5-turbo を使用した。

表 1: GPT-3.5-turbo のパラメータ

パラメータ	値
model	GPT-3.5-turbo
temperature	0.7
max_tokens	100

以下に、文から情報を抽出する際に用いたプロンプトを示す.

```
#命令
入力文を以下の形式で表現してください.
背景
時間:
場所:
登場人物の状態:
テキスト情報:
#入力文
勉強している時は、一休みすると気分が楽になる。
#条件
上記のような形式で表現してください. 背景 (時間と場所), 登場人物 (私) の状態, テキスト情報を出力してください.
時間に関して, 朝, 昼, 夜などの具体的な時間を示してください.
テキスト情報に関しては, 発言や状況などを簡潔に説明したいです. 発言がない場合は状況を適切に表現してテキスト情報を出力してください.
#例
「朝起きて、気分よくおはようと言った。」という入力
が与えられた場合,
背景
時間: 朝
場所: 自宅の寝室
登場人物の状態: 起きたばかりで気分が良い
テキスト情報: おはよう
のように出力してください. テキスト情報に関して, 上記の場合はおはようのような単語だけでお願いします.
```

以下に使用した csv ファイルの一部を示す.

表 2: 背景画像データの csv ファイル (一部抜粋)

画像 コード	画像 (場所)	画像 (時間帯, 状態)	ファイル パス
1	ATM コーナー	日中	/Users/-
2	ATM コーナー	夜	/Users/-
3	ATM コーナー	夕方	/Users/-
4	アーケード商店街	日中	/Users/-

表 3: 猫ミームデータの csv ファイル (一部抜粋)

猫ミーム コード	猫ミーム名	特徴	ファイル パス
1	DJ 猫	ひっかく	/Users/-
2	EDM 猫	EDM, 踊る	/Users/-
3	Girlfriend 猫	ポップ, 踊る	/Users/-
4	oiia 猫	ぐるぐる回る	/Users/-

## 4 実験

## 5 結果と考察

## 6 まとめと今後の展望