LLM を用いてネタバレ回避をしながらストーリー解説を提示する動画視聴支援システム

植野 天翔 * 服部 哲 * 青柳 西蔵 * 平井 辰典 *

概要. 本研究では，アニメやドラマの複雑なストーリー展開によって登場人物や内容を理解しにくくなる課題を解決するために，LLM を用いたネタバレを避けながらストーリー解説を提示する動画視聴支援システムを提案する．本システムでは，まず LLM を用いて，字幕データとあらすじ記事を入力としてタイムスタンプ付きあらすじを自動生成する．このタイムスタンプ付きあらすじをもとに，視聴中の時点までに登場した人物や場面情報を整理し，視聴中にユーザが「わからない」ボタンを押すと，LLM が現在のシーン画像および該当時刻までのあらすじに基づいて，ネタバレを含まないトピック解説と登場人物解説を生成し提示する．これにより，作品理解を補助しつつ没入感を維持する新しい動画視聴支援を実現する．

1 はじめに

アニメやドラマなどの映像コンテンツでは，登場人物の関係性や物語の背景が複雑で，視聴中に内容を把握しづらいと感じる場面や，久しぶりに視聴した際に登場人物や設定を忘れて混乱することがある．このような場合，視聴者は理解を補うためにインターネット上のストーリー解説や SNS の投稿を参照することが多いが，これらにはストーリーの先の展開が含まれている場合があり，いわゆる「ネタバレ」によって視聴体験が損なわれることがある．そこで，本研究では大規模言語モデル（LLM）を用いてネタバレを避けながらストーリー解説を生成し，視聴者に提示するシステムを提案する．

2 関連研究

ネタバレ防止に関係する研究として中村ら [3] は，スポーツ試合の結果などが含まれる Web テキストを曖昧化することでネタバレを防ぐシステムを提案しているが，既存テキストを除去・変換する受動的な手法であり，視聴状況に応じた動的な情報提示は考慮されていない．また，Uchiyama ら [2] は，授業動画の視聴時に学習者が入力した疑問に対し，LLM が即時に回答を返す学習支援システムを提案しているが，質問を能動的に入力する必要があり，視聴文脈に基づく自動的な解説生成は行われない．さらに，Kukleva ら [1] は，映画クリップを対象にキャラクター同士の相互作用と関係性を視覚・対話モダリティから予測する手法を提案しており，登場人物の動的な振る舞いや関係ネットワークを明らかにすることで映像理解を支援している．しかし，視聴者の視聴済み範囲に応じた解説生成やネタバレ制御を前提としたものではない．

本研究では，映像の再生時点までのあらすじをコンテキストとして LLM に入力し，視聴済み範囲に基づく解説を生成することで，従来のフィルタリング型手法を生成的アプローチへと拡張する．

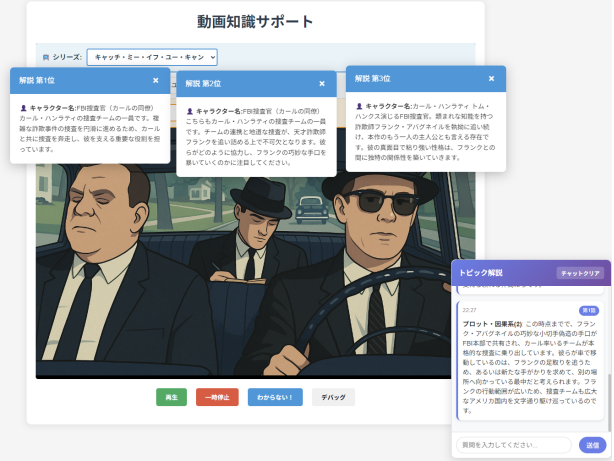

図 1. システムの UI（出典：とある映画のワンシーンを参考に生成）

3 提案手法

本研究では，ネタバレを避けたストーリー解説を生成するために，映像コンテンツの字幕やあらすじをまとめた記事を LLM へ入力する．図 1 に提案システムの UI 例を示す．視聴中にユーザが「わからない」ボタンを押すと，画面右側にはシーンに関するトピック解説が表示され，映像上には登場人物ごとの位置に対応して登場人物解説が表示される．以降では，各要素の生成手法について詳述する．

* 駒澤大学

3.1 タイムスタンプ付きあらすじの生成

映像コンテンツの字幕ファイルと既存のあらすじ記事を入力として，GPT-4o を用いてタイムスタンプ付きあらすじを自動生成する．字幕の時刻情報とあらすじ文を対応付けることで，各シーンに対応する文脈を構造化し，後述の解説生成でネタバレ範囲を制御可能とする．プロンプトでは「時刻と文を 1 行対応」「改行・空行禁止」などの形式を指定し，「00:00:00,001 → 00:00:01,263 ○○が登場する」のような出力を得るよう設計した．

3.2 登場人物情報の抽出

生成される解説の品質を高めるため，3.1 節で作成したタイムスタンプ付きあらすじから，ユーザが「わからない」ボタンを押した時点までに登場した登場人物情報を抽出する．抽出するために，GPT-4o へこれまでのあらすじデータを入力し，プロンプトにここまでの登場人物一覧を出力するように設定を行う．この情報はトピック解説および登場人物解説の両方で参照される．

3.3 トピック解説の生成

ユーザが視聴中に「わからない」ボタンを押した際，その時点のタイムスタンプとシーン画像を入力として，まずトピック解説を生成する．トピック解説では，そのシーンにおける出来事や文化的背景，専門用語などを，ネタバレを含まずに解説する．この際，LLM には表 1 に示すようなプロンプト構造を与え，利用者が抱く典型的な困惑パターンに基づいて解説を整理させる．本研究では困惑パターンを以下の 6 種類に分類し，プロンプトに用いた：

1. **キャラクター系**：登場人物の関係や立場
2. **アイテム・小道具系**：持ち物や道具の意味
3. **プロット・因果系**：展開の理由や因果関係
4. **文化・背景系**：所作や文化的文脈
5. **言語・表現系**：セリフ・比喩・ジョーク
6. **演出・技法系**：映像表現や演出意図

図 2 に，実際に生成されたトピック解説の提示例を示す．ここではテキストの入力欄が設けられており，文字を使って追加の質問などを行うことも可能になっている．

表 1. トピック解説プロンプトの構造

入力	タイムスタンプ付きあらすじ，登場人物リスト，シーン画像
指示	困惑パターン (6 分類) に基づきトピック解説を箇条書きで説明

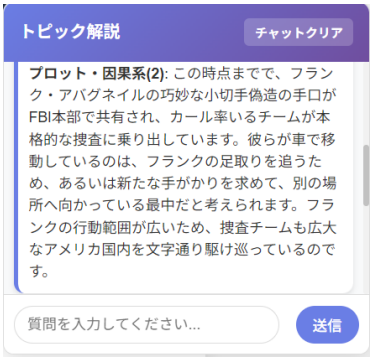

図 2. トピック解説の提示例

図 3. 登場人物解説の提示例

3.4 登場人物解説の生成

次に，ユーザが「わからない」ボタンを押した際のフレーム画像に実際に映っている登場人物のみを対象に，位置情報付きで登場人物解説を生成する．本解説では，表 2 のような構造のプロンプトを与え，LLM のマルチモーダル機能を用いて画像解析を行い，登場人物の特定と説明を同時に行う．図 3 に登場人物解説の提示例を示す．出力された位置情報を元に，登場人物ごとに吹き出し形式で解説が重ねて表示され，それぞれの人物についての解説を確認をすることができる．

4 まとめ

本研究では，アニメやドラマなどの映像コンテンツ視聴時に，ネタバレを避けながらストーリー理解を支援するための動画視聴支援システムを提案した．

今後の課題として，本システムについての評価実験や各ユーザに最適化された解説の提示，GPT-4o 以外の LLM モデルによる性能比較を検討している．

表 2. 登場人物解説プロンプトの構造

入力	タイムスタンプ付きあらすじ，登場人物リスト，シーン画像
指示	映っている登場人物のみを特定し，位置座標とともに登場人物についての解説を出力

参考文献

[1] A. Kukleva, M. Tapaswi, and I. Laptev. Learning interactions and relationships between movie characters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9849–9858, 2020.

[2] S. Uchiyama, K. Umemura, and Y. Morita. Large Language Model-based System to Provide Immediate Feedback to Students in Flipped Classroom Preparation Learning, 2023.

[3] 中村聡史, 小松孝徳. スポーツの勝敗にまつわるネタバレ防止手法: 情報曖昧化の可能性. 情報処理学会論文誌, 54(4):1402–1412, 2013.

未来ビジョン

自身の経験として，アニメやドラマを視聴しているときに,「この人誰だっけ？」と感じて，一度停止して登場人物について調べてしまい，そのままネタバレを見てしまった経験がある．せっかく作品の世界に没入していたのに，先の展開を知ってしまったことで，その後の楽しみや緊張感が薄れてしまう体験を何度も経験してきた．同じように，内容を理解したいのにネタバレのリスクが怖くて調べられない，というジレンマを感じたことのある人は多いのではないだろうか．

本研究では，この「理解したいのに調べられない」という体験を解消し，作品への没入感を保ちながら理解を深められる視聴体験を実現することを目指している．将来的には,「わからない」ボタンを押す必要すらなく，視聴者が混乱しそうな場面を自動で検出し，適切なタイミングでストーリー解説を提示できるシステムを目指す．視聴者の視線，反応，過去の視聴履歴などをもとに，個人ごとに「理解が追いついていない」兆候を推定し，その人に最適化された登場人物やトピックの解説をリアルタイムに提示する．

さらに，将来的には複数人での映画・アニメ視聴にも対応し，それぞれの視聴者が「自分にだけ必要な情報」を VR ヘッドセットや AR グラスを通して確認できるような環境を想定している．例えば，ある登場人物を久しぶりに見る視聴者にはその人物の説明が自動的に表示される一方で，すでに理解している別の視聴者には何も表示されない．このように，視聴者一人ひとりの理解度と興味に応じて情報提示を最適化することで，没入感を損なわずに物語理解を支援する視聴体験の実現を目指す．