# 特徴タグ分析を用いたアニメシーンに関するメタデータ作成の提案

第34回 インタラクティブ情報アクセスと可視化マイニング

SHAN Junjie, 石井智也, 安尾萌, 西原陽子

立命館大学情報理工学部
立命館グローバル・イノベーション研究機構

2025-03-07

RITSUMEIKAN

# 目次

- 背景&問題点
- 解決したい内容
- 提案システム
  - カットシーンの分割
  - フレーム画面の内容タグの検出
  - カットシーンの特徴タグの集約
- 分析
  - ジャンル別のアニメのカットシーン
  - ジャンル別アニメの頻出タグ
- おわりに

# 背景&問題点

- LLMの発展により入力ビデオの説明文の作成は簡単になる
  - 説明文に通じて具体的な応用（分類・検索等）にも至る

- LLMの特性と自然言語自身に柔軟性があり、記述された説明文は統一性を保つのは困難
  - 統一的な分析・インデクスの構築は不可

+

Prompt：
このシーンの説明文を書いてください

緑豊かな公園の遊具のそばで、三人の少女たちが仲睦まじく遊んでいる。しかし、そのすぐ横で、一人の少女が膝を抱えて寂しそうに座り込んでいる。彼女の表情は暗く、何か悲しいことがあったかのようだ。三人組とにこやかな笑顔とは対照的に、彼女の孤独感が際立っている。

画像には、以下の要素が含まれているようです。
背景:緑豊かな公園の風景が...木々、遊具...
人物: 左側に3人の少女が...互いに腕を...
　　右側に1人の少女が座り込んでおり...
服装: 3人の少女は、同じ紺色のスカート...
　　座り込んでいる少女は、白いシャツに...
その他: 空には白い雲が浮かんでいます。

# 解決したい内容

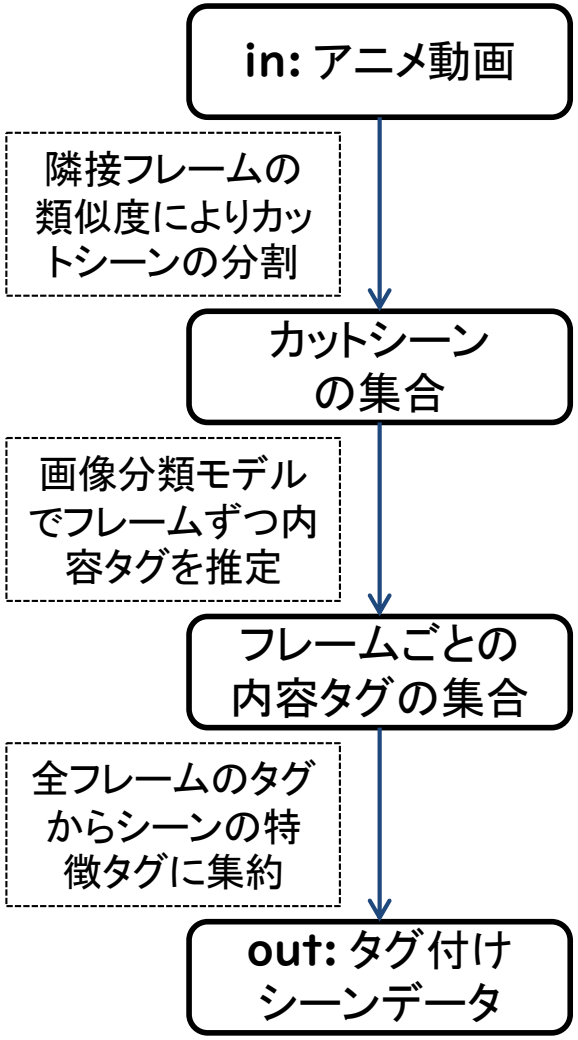
1. アニメのビデオコンテンツに対する、画面の内容を反映できる統一的な説明記述（特徴タグ）を作成する

2. 特徴タグを種類に分けて、比較可能なアニメのシーンデータの内容を記述するメタデータの自動作成

| | |
|---|---|
| Scene_01 | 1boy, 2girls, black_hair, blue_eye, smile, … outdoor, sky, … upper_body, eye_focus … |
| Scene_02 | 1boy, 1girl, brown_hair, black_eye, shirt, … indoors, window, … from_above |
| Scene_03 | 2boys, 2girls, white_hair, purple_eye, shout, … building, city, … depth_of_field |

# 提案手法の流れ

- 入力：アニメの動画ファイル
- カットシーンの分割：
  - 隣接フレーム間の類似度の比較
  - ミスした分割に対応するルール設定
- カットシーンに対する特徴タグ：
  - フレームごとの内容タグの推定
  - カットシーンの内容を代表する特徴タグの集約
- 出力：特徴タグの付けたシーンデータ



# アニメのカットシーン

- 「カットシーン」は、動画構成の最小単位で、一つの構図を表現し、同一キャラやオブジェクトを表現する連続のフレームの集合である。
  - ビデオコンテンツでは、画面内容の一貫性のある部分

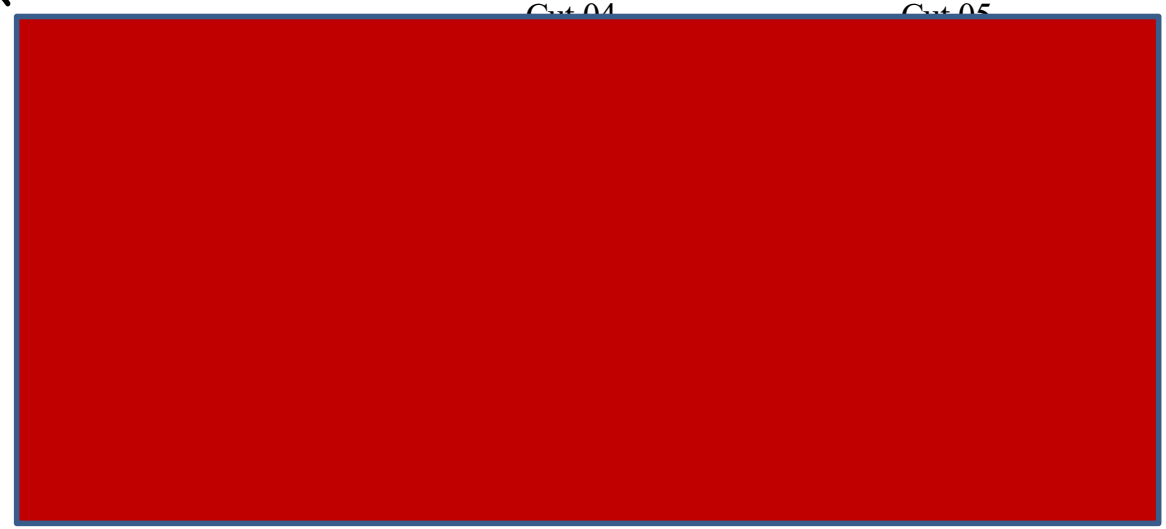
https://orita-ani.net/scene-different-from-sequence/

Scene 01

# アニメのカットシーン

- 「カットシーン」は、動画構成の最小単位で、一つの構図を表現し、同一キャラやオブジェクトを表現する連続のフレームの集合である。
  - ビデオコンテンツでは、画面内容の一貫性のある部分

https://orita-ani.net/scene-different-from-sequence/



Scene 01

# カットシーンの分割

- フレームごとに隣接フレーム間の類似度を計算する、類似度が低い隣接フレームの前後に分ける。
  - 使用した類似度はOpenCVのテンプレートマッチングにある「TM_CCOEFF_NORMED」である（−1 ～ +1）。
  - 設定した類似度「TM_CCOEFF_NORMED」の閾値は「0.6」となる。

    隣接フレーム間の類似度は0.6未満の場合は、隣接フレームの前後に分けて、前のフレームまでの連続するフレーム集合を一つのカットシーンとして分割する。
  - フレーム数が10枚未満のカットシーンが連続に出現した場合、順番に合わせて前からフレーム数10枚未満のカットシーンを合併する（≧30枚まで）。

- 人手でランダムの3000件の分割されたカットシーンをチェックした結果は、分割精度は93.83%となった。

# フレーム画面の内容タグの推定

- 分割されたカットシーンに対して、フレームごとに画面内容の説明記述（内容タグ）を推定する
  - WEB上の公開モデルを使用[1]
    （画像をマルチラベルに分類するモデル、1ラベル→1タグ）
  - 推定可能の内容タグ数は合計6891個
  - 本研究では、その中にある3種類のタグを使用
    Character（キャラ関連）、
    Background（背景関連）、
    Composition（構図関連）
  - 3種類で合計3264個

| タグ種類 | サンプル | タグ説明 |
|---|---|---|
| Character<br>キャラの関連タグ<br>計2748個 | blue_eye | 青い目のキャラがいる |
| | blonde_hair | 金髪のキャラがいる |
| | shouting | 叫んでいるキャラがいる |
| Background<br>背景の関連タグ<br>計394個 | cloud | 雲が背景にある |
| | classroom | 教室の風景がある |
| | indoors | 室内である背景 |
| Composition<br>構図の関連タグ<br>計122個 | from_above | ハイアングル、俯瞰 |
| | pov | 一人称視点 |
| | eye_focus | 目のクローズアップ |

1. https://huggingface.co/spaces/hysts/DeepDanbooru

# カットシーンの特徴タグの集約

- フレームごとに推定した内容タグを図1に示した形で記録
  - カットシーンのID
  - カットシーンの合計フレーム数
  - 内容タグの格納ベクトル（各内容タグがカットシーンでの出現回数を記録）

```
CaptainTsubasa_005_cut0281.mp4:
[44, [0, 0, 44, 0, 0, ..., 0, 5, 0, 0, 0]]
カットシーンのID
Jujutsu_004_cut0049.mp4:
[55, [0, 3, 11, 0, 0, ..., 0, 0, 0, 0, 0]]

UmaMusume_S1_005_cut0065.mp4:
[226, [0, 0, 0, 0, 0, ..., 0, 70, 0, 0, 0]]
カットシーンの合計フレーム数   当該タグの出現したフレーム数
HeroAcademia_010_cut0329.mp4:
[36, [0, 22, 0, 0, 0, ..., 0, 0, 0, 0, 0]]
長さ6891の内容タグの位置に対応する格納ベクトル
HibikeEuphonium_S1_004_cut0334.mp4:
[56, [0, 56, 56, 0, 0, ..., 0, 0, 0, 0, 0]]
```

図1

- 各カットシーンに対して、カットシーンの合計フレーム数の50%以上の出現頻度のタグを、当該カットシーンの特徴タグとして残ってシーンに付与する
  - 同じカットシーンにしても、画面の内容も常に変わっている
  - カットシーンの特徴タグをOne-Hotベクトルに保存する

```
HeroAcademia_005_cut0073.mp4 [0, 0, 0, 0, 0, ... 1, 0, 0, 0, 0]
MIX_S1_001_cut0093.mp4       [0, 1, 1, 0, 0, ... 0, 1, 0, 0, 0]
DragonBall_107_cut0275.mp4   [0, 1, 1, 0, 0, ... 1, 0, 0, 0, 0]
```

長さ6891のOne-Hotベクトル

# タグ付けシーンデータに対する分析

- アクション・スポーツ・日常系の3つジャンルからアニメ動画を300話収集し、提案手法により特徴タグ付けのシーンデータを作成した
  - 合計99138本のカットシーン、平均フレーム数は122.94枚

- 内容タグがアニメのカットシーン内での出現分布を調査した
  - タグの出現頻度に分けるタグ数の割合
  - 50%以上(39.34%+11.16%)の内容タグがカットシーン内で20%未満のフレームにしか検出されなかった
  - 18%以上(5.56%+12.62%)の内容タグが90%以上のフレームに存在する

表1：ジャンル別のカットシーン数

| ジャンル | カットシーン数 | 平均フレーム数 |
|---|---|---|
| アクション | 34673 | 127.97 |
| スポーツ | 31864 | 119.91 |
| 日常系 | 32601 | 120.54 |
| 合計 | 99138 | 122.94 |

表2：内容タグがカットシーンでの出現頻度

| 出現頻度 | タグ数割合 |
|---|---|
| 10%未満 | 0.3934 |
| 10% ～ 20%未満 | 0.1116 |
| 20% ～ 30%未満 | 0.0723 |
| 30% ～ 40%未満 | 0.0552 |
| 40% ～ 50%未満 | 0.0446 |
| 50% ～ 60%未満 | 0.0401 |
| 60% ～ 70%未満 | 0.0352 |
| 70% ～ 80%未満 | 0.0320 |
| 80% ～ 90%未満 | 0.0337 |
| 90% ～ 100%未満 | 0.0556 |
| =100% | 0.1262 |

# ジャンル別の頻出タグ

- ジャンル・タグ種類別の頻出タグTop5

| タグ種類 | アクション | | スポーツ | | 日常系 | |
|---|---|---|---|---|---|---|
| | タグ内容 | 出現シーン数(割合) | タグ内容 | 出現シーン数(割合) | タグ内容 | 出現シーン数(割合) |
| Character | 1girl | 11228 (32.38%) | black_hair | 12875 (40.41%) | 1girl | 13881 (42.58%) |
| | 1boy | 8722 (25.15%) | shirt | 10695 (33.56%) | long_hair | 13117 (40.23%) |
| | black_hair | 8405 (24.24%) | 1girl | 9281 (29.13%) | brown_hair | 10490 (32.18%) |
| | short_hair | 5353 (15.44%) | 1boy | 9159 (28.74%) | short_hair | 9327 (28.61%) |
| | open_mouth | 5344 (15.41%) | brown_hair | 7626 (23.93%) | open_mouth | 7418 (22.75%) |
| Background | sky | 8205 (23.66%) | outdoors | 8640 (27.12%) | outdoors | 6590 (20.21%) |
| | outdoors | 7813 (22.53%) | sky | 5917 (18.57%) | sky | 5069 (15.55%) |
| | cloud | 6824 (19.68%) | blurry_bg | 4450 (13.97%) | blurry_bg | 4308 (13.21%) |
| | blue_sky | 4264 (12.30%) | tree | 4426 (13.89%) | indoors | 4271 (13.10%) |
| | cloudy_sky | 4187 (12.07%) | cloud | 4074 (12.79%) | tree | 4117 (12.63%) |
| Composition | male_focus | 9034 (26.05%) | male_focus | 11258 (35.33%) | depth_of_field | 5515 (16.92%) |
| | close-up | 3897 (11.24%) | upper_body | 6032 (18.93%) | upper_body | 3293 (10.10%) |
| | upper_body | 3533 (10.19%) | depth_of_field | 5205 (16.34%) | close-up | 2440 (7.48%) |
| | depth_of_field | 2276 (6.56%) | close-up | 2729 (8.56%) | male_focus | 1681 (5.16%) |
| | from_behind | 567 (1.63%) | from_side | 871 (2.73%) | gradient | 674 (2.07%) |

- アニメ動画はジャンルに関係なく、室外や自然的景色がある場所のシーンが多い
- アニメ動画では、上半身(upper_body)を描写するシーンが多い
- 「アクション」アニメでは、近距離からの描写(close-up, 11.24%)が多い、「スポーツ」と「日常系」アニメでは、被写界深度（depth_of_field）の表現が多い（16.34%，16.92%）

# おわりに

- 画像内容により説明タグの推定モデルを利用し、アニメ動画のカットシーン向きの特徴タグのデータ構築手法を提案した
- 三つのアニメジャンルから、提案手法により特徴タグ付けのシーンデータを作成した
- 作成したアニメシーンのタグデータを分析し、アニメジャンルでの作画の傾向性を考察した

- アニメシーンのメタデータ構築の完成に2つの課題がある：
  - タグ間に上位・下位関係がある、細かく分類する必要
  - タグに対応する画面の部分（位置情報）の判明・抽出