

動画のハイライト部分自動抽出に向けた検討

小城, 凱
九州大学大学院システム情報科学府

伊東, 栄典
九州大学情報基盤研究開発センター

<https://hdl.handle.net/2324/7157274>

出版情報 : FIT : forum on information technology. 22 (第2分冊), pp.473-474, 2023. Information Processing Society of Japan

バージョン :

権利関係 : 本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。



動画のハイライト部分自動抽出に向けた検討 A study of relationship between video highlights, user chats and genres

小城 凱¹⁾ 伊東栄典²⁾
Kai Koji Eisuke Ito

1 はじめに

近年、YouTube を始めとする投稿型動画サイトが人気で、膨大な数の動画が毎日投稿されている。ライブ配信も人気で、視聴者の多い配信者は定期的にライブ動画配信を行っている。ライブ配信動画は、ライブ（同時刻）視聴だけでなく、YouTube の機能を活かしての後から視聴も多い。

2010 年代のスマートフォン普及により動画視聴の短時間化の需要が高まる。稲田は著書 [1] の中で、時間効率向上のため映画を早送りで見ることが増えたと示している。時間短縮需要に応じて TikTok やショート動画が流行している。2010 年代には長時間動画の一部を短く切抜いた動画（以降、切抜き動画）の投稿が現れた。その後、切抜き動画の投稿者と、元動画投稿者の両方が利益を得る仕組みができ、切抜き動画が増加した。

我々は機械学習による、動画のハイライト部分および要約部分の抽出を目指している。ハイライト抽出のための機械学習データとして切抜き動画を考えている。切抜き前の元動画の、ハイライト部分が切抜き動画であると仮定する。切抜き動画に基づいて、元動画の切抜き部分を機械学習で抽出できれば、同種類の動画に対するハイライト部分の自動抽出が可能になる。

本論文では、Vtuber のゲーム実況動画を対象に、切り抜かれる前のライブ配信動画と、切り抜かれた切り抜き動画を比較する。ハイライト部分は、元のライブ配信動画における視聴者のチャットやコメントの増加する時間帯ではないかと仮定する。チャットの盛り上がり部分が、動画ハイライトとして切抜き動画に用いられているかを検証する。

本論文の構成は以下の通りである。第2節で関連研究を述べる。第3節では、本論文の分析対象にした動画について述べる。第4で、切抜かれる元動画への視聴者チャット状況と、切抜き動画との関係分析を示す。最後に第5で、まとめと今後の課題を述べる。

2 関連研究

動画のハイライト部分抽出について、動画内容を対象とするコンテンツベースの手法が提案されている。小林らは、映画の予告編やスポーツのハイライトシーンから、元動画の同じシーンを探す手法を提案している [2]。具体的には深層強化学習による効率的な動画内類似画像検索手法を提案している。予告編やハイライト動画の代表的シーンをクエリ画像とし、フレームスキップ行動を強化学習する。クエリ画像と現在フレームの画像特徴ベクトル間の距離（特徴量距離）を考慮した特徴抽出モジュール、報酬設計の工夫により、精度を向上させた。実験の結果、液体注ぎ動画のデータセットである Multiview Pouring Dataset において、検索成功率 83.0%

と述べている。小林らの手法は、クエリ画像のシーンを動画から検索する手法で、本研究が対象とする切抜き動画のような「視聴者が好む部分の自動抽出」とは異なる。

Wang らは人手による注釈無しに長編映画のハイライトシーン抽出について研究している [3]。映画の予告編 (trailer) は映画本編のハイライトシーンであると仮定し、予告編を機械学習して長編映画から重要瞬間を検出するモデルの構築を試みている。この手法はコンテンツベースのハイライトシーン抽出である。本論文で我々が検討するのは、視聴者コメントに基づく（ソーシャルベースの）ハイライトシーン抽出である。

Muszynski らは映画のシーンが観客に与える感情について分析している [4]。彼らは、感情的な映画について、認識（知覚感情）と観客が引き起こされる感情（誘発感情）の間の関係を研究している。知覚感情と誘発感情の関係分析のため、彼らはクラウドソーシング方式で映画に知覚感情の注釈を付けている。この注釈情報を用いて、観客の生理学的反応および行動的反応、知覚感情、映画内の会話、美的ハイライトが、映画によって引き起こされる誘発感情を区別するものであることを示している。Muszynski らの研究は、視聴者の感覚を調べる点で、本研究に近い。一方、本研究では観客の感情値に基づくハイライトシーン抽出を検討するため、因果関係が逆である。

3 分析対象動画

先に述べた様に、我々は Vtuber のゲーム実況動画を対象に、切り抜かれる前のライブ配信動画と、切り抜かれた切り抜き動画を比較する。元のライブ配信動画のハイライト部分が切抜き動画であるとし、ライブ配信動画で視聴者のチャットやコメントが盛り上がる時間帯が、切り抜かれる部分に当たるのかを調査した。

この予備調査のため、表1に示す8個の動画に対して小規模に調査した。表に示す動画はいずれも Vtuber によるゲーム実況配信動画である。表1の user 数は動画にチャットやコメントを投稿した利用者 ID の数で、chat 数は投稿されたチャットの総数である。

表1 予備調査対象動画

no	video id	公開日	動画長	user 数	chat 数
1	FupqygeIUrK	2023/04/10	2:02:25	6828	53,772
2	lzqDVQTKBEI	2022/12/25	1:20:01	2951	32,282
3	2SOUuyVJnTg	2023/03/11	1:29:49	347	5,255
4	6TmEPJAqopY	2023/03/26	3:20:39	6508	53,220
5	ZhhZmmTN9NM	2022/12/27	1:14:28	356	4,019
6	jdXIF2mezuw	2023/02/21	2:55:41	801	6,473
7	kptxnF6iR-g	2022/12/25	1:19:35	1602	19,131
8	nv9RiHGy_fs	2023/04/14	3:41:48	4615	33,569

4 元動画のチャットと切抜きの動画の関係

表1に示す8つのライブ配信動画を対象に、元のライブ配信動画で視聴者が盛り上がる部分が切り抜かれてい

1) 九州大学大学院システム情報科学府
2) 九州大学情報基盤研究開発センター

るかを分析する。YouTube API¹⁾を用いて、8つの動画のメタデータと視聴者チャットを入手した。ライブ配信動画での視聴者チャットの盛り上がり、と、切抜き部分との関係を調査した。切抜き動画は、一人の切抜き動画投稿者の動画のみを対象に分析している。

図1に、ライブ配信動画に対するチャット投稿数の時系列変化（折れ線）と、切り抜かれた部分（オレンジ色）との関係を示す。各グラフの動画は、表1の順序で示している。チャット投稿数の時系列変化は、平滑化のために5分の移動平均線で描画している。

図1を見ると、チャットが盛り上がる時間帯を切り抜いている事が分かる。しかしながら、チャット投稿数の最大盛り上がり部分とは言えない。動画の前半、中間、後半についても切抜き部分が偏っていない。

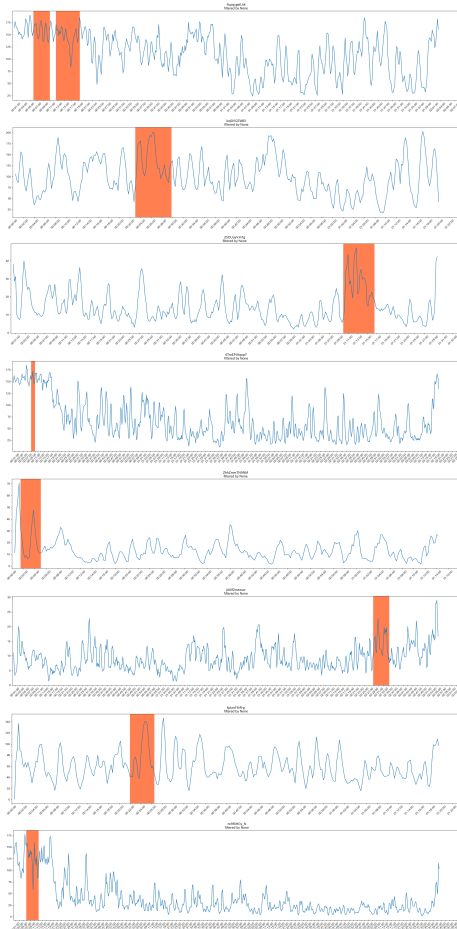


図1 時系列でのチャット投稿数と切抜き部分

次に視聴者が動画を楽しんでいる状況だけを調べるため、「笑い」に関する文字列を含むチャットだけに限定した。「笑い」に関する文字列は、「笑, ww, 草」の3つを含むものとした。図2に結果を示す。図2を見ると、動画1,2,3,4は笑いの単語が多い部分が切抜き部分に重なる。しかし動画5,7,8はそうではない。動画6は笑いの単語がチャット文に含まれず、分かりやすい結果を得ていない。

1) <https://developers.google.com/youtube/v3?hl=ja>



図2 笑い限定チャットと切抜き部分

5 おわりに

本論文では、8つのライブ配信動画を対象に、ライブ配信動画での視聴者チャットの盛り上がり、と切抜き部分との関係を調査した。その結果、概ねチャット投稿数が盛り上がる部分が切り取り対象であることがわかった。また笑いに関するチャットに限定した場合も同等の結果が得られた。

今後は更に詳細な分析を行う。チャット投稿が多い人だけに限定する場合や、チャット投稿が多すぎる人の影響を和らげた分析を行う。他にも対象動画を増やしての分析を行いたい。動画内容について、特に音声分析を組み込みたい。

参考文献

- [1] 稲田豊史, 映画を早送りで見ると人々, 光文社 (2022).
- [2] 小林弘人 et.al., 動画内類似画像検索のためのフレームスキップ行動の深層強化学習, 画像電子学会研究会講演予稿, vol.21.03, no.0, pp.88-93 (2022).
- [3] Lezi Wang, et.al., Learning Trailer Moments in Full-Length Movies with Co-Contrastive Attention, Computer Vision – ECCV 2020, pp.300-316 (2020).
- [4] Michal Muszynski, et.al., Recognizing Induced Emotions of Movie Audiences from Multimodal Information, IEEE Transactions on Affective Computing, vol.12, no.1, pp.36-52 (2021).