

CPEX：配信動画の視聴者投稿コメントに基づく 切り抜き動画の自動制作に関する検討

松田 隼弥¹ 大井 翔¹

概要：現在、インターネット上で動画配信（ライブ配信）を行う人が増加している。また、動画投稿サイトに投稿される動画や配信の時間も1時間を超えるような動画も増えてきている。このような背景の中、長時間の配信動画を見やすくして提供する切り抜き動画などがあり流行している。切り抜き動画のような本来長い動画を短くし効率よくコンテンツとして消化するのは今の若者にあっているものだと考えられる。またそのためか youtube では Short 動画という 10～60 秒ほどの短い動画の比率も上がってきている。その中で切り抜き動画を作る人は増えつつあるが、配信直後の切り抜きなどはまだ作られていない場合が多いので配信を見逃してすぐにその配信の面白い部分だけを見ることはできない。その問題を解決するためにライブ配信のコメントから自動的に切り抜き動画を誰でも作れるようにするシステムを作る。この研究の目的はシステムで制作した切り抜き動画が既存の切り抜き動画と比較した時に、どれだけ内容をまとめた切り抜き動画として見やすく分かりやすいものになっているか調査することである。

1. はじめに

近年、動画共有サービスの普及と発展に伴い、ユーザが動画共有サービスの利用に使う時間が長くなっている [1]。その中で図 1 に示すように youtube では Short 動画や切り抜き動画のような短い動画の投稿比率が上がっている [2]。

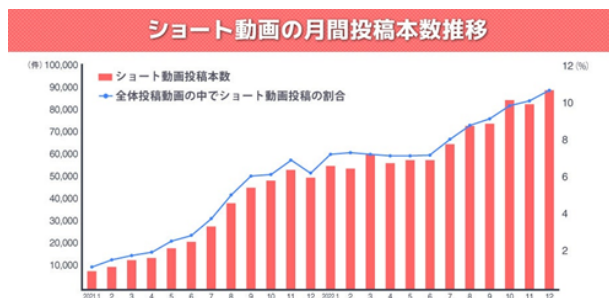


図 1 ショート動画の増加推移 [2]

これを受けてユーザが効率的に動画共有サービスを利用し、目的の動画をより短時間で視聴できる手法が必要とされていると我々は考える。しかし切り抜き動画というものは切り抜き作業があるため、ライブ配信直後にその配信の切り抜き動画は存在しない。また2時間の動画を10分の動画にまとめる動画の編集には基本的に5時間～20時間かかる [3] ため切り抜き動画の作成も同じだけの時間がかか

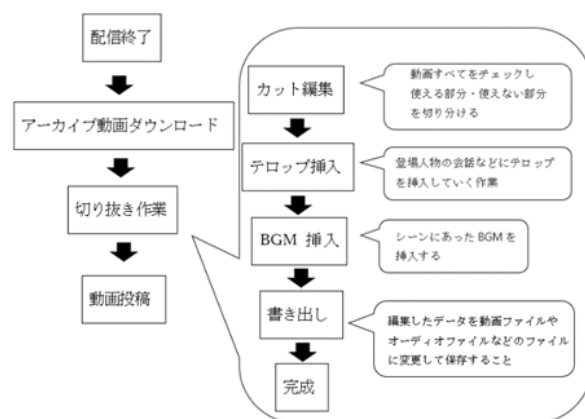


図 2 切り抜き動画編集の主な作業工程

ると考えられる。図 2 は切り抜き動画編集の主な作業工程を示す。

本研究はライブ配信直後の切り抜きが存在しない時でも配信動画の切り抜きが誰でも見られるようにするためにアーカイブ動画の時刻同期コメントから自動でどのようなシーンなのか判定しそのシーン別で切り抜き動画の作成を行う。どのようなシーンであるかの定義付けとして youtube の様々な配信動画で出現頻度の高いコメントがどの程度の感情を込められて打たれたものなのか調査し、そのコメントに含まれた感情の要素がそのシーンにどれだけ含まれているかを明らかにしたうえでそのシーンがどんなシーンなのか判別する。その判別したシーンの中で「面白い」と

¹ 大阪工業大学

いった感情が多く含まれているシーンは「面白いシーン」と判別し、そのシーンのみの切り抜き動画を作成する。また判別するシーンを選定するときにコメント数が少ない箇所は感情の要素数も少なく、ほんとにその感情のシーンであるのか断定しづらいため、コメント数の多いシーンから判別を行う。

本研究では図3のシステムのフロー図の工程である時刻同期コメントから自動でどのようなシーンなのか判別し切り抜きシステムを参加者に使用してもらい評価してもらう。その評価をしてもらう過程で比較対象として既存の人の手によって作られた切り抜き動画を用いて、どちらのほうが切り抜き動画として見やすいものなのかを評価してもらう。そしてシステムが既存の切り抜き動画と遜色のないレベルで見やすく、どのような切り抜き動画かわかるものを作成できるようにすることが目的である。

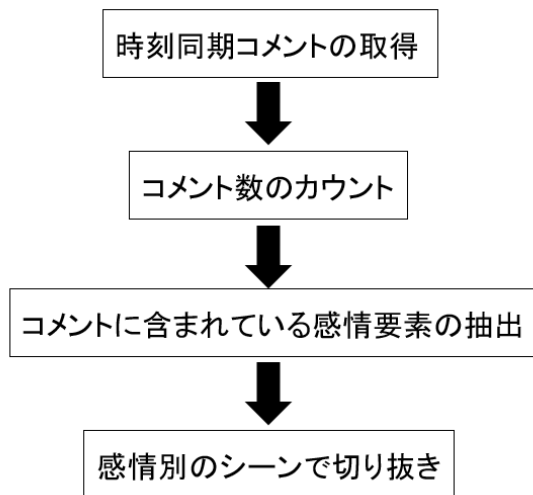


図3 提案システムのフロー図

2. 関連研究

本章では、ユーザのコメントを用いた研究を本研究に近いものとして挙げる。またユーザのコメントを対象とした研究は、シーン分析(2.1節)とコメント分析(2.2節)とに分けて考えられるので、各節でそれぞれについて述べる。

2.1 シーン分析

末永ら[4]はニコニコ動画の時刻同期コメントから自動で感情分類を行い、コメントが多く投稿されている箇所を特徴シーンとし同じ内容のコメントの中でその数が多かったコメントを特徴コメントとし、実際に動画を見て評価した評価者が特徴シーン、特徴コメントと推定したものとどの程度一致するかの研究を行っていた。

2.2 コメント分析

鳥海ら[5]は2019年から流行したコロナウイルスに関

するTwitterのツイートコメントを収集し、感情分析ツールML-Askを用いてコメントの感情を(哀, 恥, 怒, 厭, 怖, 驚, 好, 昂, 安, 喜)の10感情で判定してどのような感情のツイートコメントが多いのかを調査した。この2つの研究ではシーンの分析とコメントの分析を行っており、本研究でこの要素を組み合わせる切り抜き動画を作成し、それがシーン別の切り抜き動画となっているのか調べる。

3. 対象となるサービス

Youtubeは2023年3月の時点でアクティブユーザーが全世界で25億1400万人存在し、日本では7000万人以上存在している動画共有サービスである[6]。Youtubeでは2011年4月からYoutubeライブというライブストリーミングサービスが開始し、リアルタイムでのコメントの投稿が可能となった。また配信後にはアーカイブ動画として動画の再生時間と同期したコメントがチャットリプレイで見られる仕組みがある。このためアーカイブ動画ではコメントの投稿が多い再生箇所に視聴者が注目していることが分かる。このシーンを見どころシーンとして扱い、切り抜き箇所として切り抜き動画を作成する。また対象とするアーカイブ動画はゲーム実況配信とする。これは図4に示すように現在Youtubeの動画でゲーム配信の割合が多く需要が高いと考えられるからである。

10代 (10代)		20代 (20代)		30代 (30代)		40代 (40代)					
1	音楽 (MV/PV/バラ)	52.2%	1	音楽 (MV/PV/バラ)	38.9%	1	音楽 (MV/PV/バラ)	37.1%	1	音楽 (MV/PV/バラ)	35.7%
2	ゲーム	47.9%	2	ゲーム	36.5%	2	面白い / パラエディター	24.1%	2	面白い / パラエディター	22.0%
3	マンガ / アニメ	24.9%	3	面白い / パラエディター	28.4%	3	ゲーム	21.7%	3	料理 / グルメ	17.4%
4	アイドル / 芸能人	24.1%	4	アイドル / 芸能人	23.7%	4	アイドル / 芸能人	16.0%	4	ペット / 動物	16.7%
5	面白い / パラエディター	22.1%	5	化粧品 / コスメ	16.9%	5	料理 / グルメ	15.7%	5	スポーツ	15.2%
6	チャレンジ / やって見た	21.0%	6	料理 / グルメ	16.5%	6	運動 / フィットネス (トレーニング/ヨガ含む)	14.2%	6	アイドル / 芸能人	14.5%
7	化粧品 / コスメ	18.7%	7	スポーツ	15.0%	7	スポーツ	13.9%	7	ゲーム	13.9%
8	雑談	17.8%	8	チャレンジ / やって見た	14.6%	8	マンガ / アニメ	12.7%	8	乗り物 (車/バイク/電車など)	12.2%
9	教育 / 勉強	17.0%	9	マンガ / アニメ	14.6%	9	ペット / 動物	10.8%	9	運動 / フィットネス (トレーニング/ヨガ含む)	11.9%
10	Vtuber (バーチャルYouTuber)	15.1%	10	紹介 / レビュー	13.8%	10	教育 / 勉強	9.8%	10	マンガ / アニメ	11.0%

図4 年代別のよく見るジャンルの動画 [7]

4. CPeX (Clipping Excitement) システム

提案手法としては動画内の時刻同期コメントをYoutubeのアーカイブ動画から取得し、コメントが多く投稿されているシーンを1番コメントが多いシーンから順に30個~60個推定し、そのシーンを10秒で自動で切り抜き、切り抜き動画を作成するものである。また切り抜きシーンを10秒にしたのは人間の初頭効果という与えられた情報がその後の情報にも影響し興味を引きやすくなる秒数だからである[8]。また30個~60個の切り抜きシーンにし、動画全体の長さを5分にしたのは図5の週3以上切り抜き動画を見ている人の見ている切り抜き動画の長さが5~10分であるためである。また各コメント数の多いシーンのコメントか

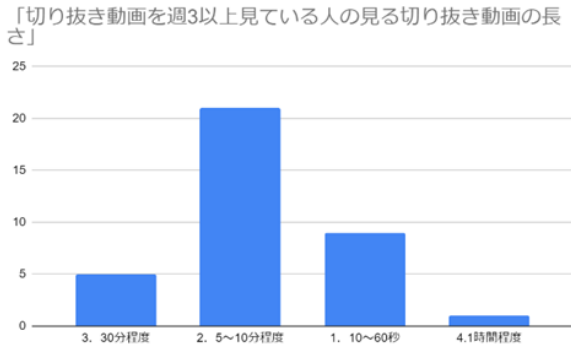


図 5 週 3 以上切り抜き動画を見ている人の見ている切り抜き動画の長さ

ら感情要素を振り分けした単語が含まれている場合、そのシーンの感情の判定を行い、感情別に 1 番多い感情要素のシーンを 1 番から順に 30 個推定し、そのシーンを 3~10 秒で自動で切り抜きを行い感情別切り抜きを作成する。図 6 と図 7 にその概要を示す。

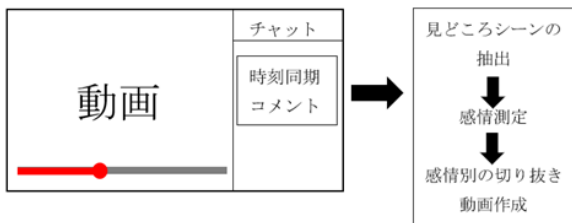


図 6 提案手法の概要図

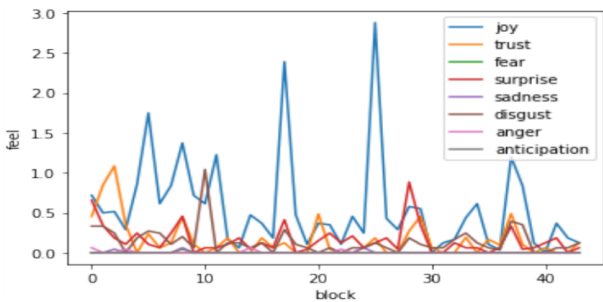


図 7 感情別シーンのイメージ図 [9]

この感情別の切り抜きをユーザが選択できるようにし、自分の見たい切り抜きを作れるようにすることが現在のシステムの目標である。

5. 実験

本研究では Cpex でコメントを感情別に振り分けて感情要素を取得するために実験として 20 人の大学生を対象としてネットでよく使用されている 53 個の単語を動画配信でコメントとして打った時を想定して「楽しい、信頼、恐怖、驚き、悲しい、嫌悪、怒り、期待」の 8 つの感情で 1~10 段階で評価してもらった。これは感性モデルとして、プ

表 1 単語と感情ごとの確率

	楽しい	信頼	恐怖	驚き	悲しい	嫌悪	怒り	期待
!	0.19	0.13	0.09	0.26	0.05	0.06	0.07	0.15
w	0.33	0.16	0.06	0.13	0.05	0.07	0.07	0.14
ゴミ	0.06	0.05	0.05	0.05	0.06	0.37	0.30	0.05
ざまあ	0.19	0.08	0.05	0.07	0.06	0.24	0.20	0.12

ルチックの感情の輪モデルを参考にした [10]. 下に図 8 に実験アンケート画面を示す。

ネットスラングコメントの感情の幅の調査

「えええ」というコメントをゲーム配信などで打った場合それにはどの感情がどの程度含まれていますか？

楽しい ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 信頼 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 恐怖 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 驚き ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 悲しみ ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 嫌悪 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 怒り ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10
 期待 ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 ○10

図 8 実験のアンケート画面

これらのデータを収集してそれぞれの単語に対して取得した感情要素を感情ごとに合計した。また単語ごとにも取得した数値を合計した。下の図 9 にその値を示す。

文字	楽しい	信頼	恐怖	驚き	悲しい	嫌悪	怒り	期待	合計
...	105.0	71.0	47.0	143.0	29.0	33.0	39.0	84.0	561.0
...	60.0	53.0	37.0	62.0	23.0	37.0	46.0	61.0	387.0
...	29.0	33.0	33.0	40.0	54.0	43.0	59.0	31.0	322.0
...	28.0	30.0	36.0	44.0	61.0	41.0	65.0	31.0	336.0
w	224.0	110.0	38.0	87.0	32.0	47.0	44.0	91.0	673.0
wsk	93.0	91.0	20.0	25.0	19.0	19.0	21.0	111.0	358.0
ww (wの複数形)	226.0	109.0	38.0	84.0	38.0	49.0	53.0	91.0	688.0
おおお	64.0	46.0	42.0	65.0	64.0	40.0	53.0	32.0	436.0
しいい	44.0	38.0	29.0	33.0	27.0	31.0	34.0	27.0	263.0
ううう	37.0	30.0	27.0	35.0	52.0	38.0	47.0	26.0	292.0
うま	126.0	116.0	33.0	121.0	27.0	27.0	27.0	108.0	565.0
えええ	49.0	41.0	38.0	79.0	32.0	38.0	46.0	36.0	359.0
おおお	74.0	55.0	36.0	82.0	28.0	28.0	27.0	75.0	425.0
おもしろ	175.0	98.0	35.0	49.0	31.0	36.0	31.0	84.0	538.0
か	29.0	20.0	30.0	25.0	35.0	145.0	137.0	21.0	442.0
えええ	93.0	91.0	25.0	28.0	19.0	19.0	19.0	111.0	358.0
ゴミ	25.0	20.0	21.0	20.0	23.0	144.0	119.0	21.0	393.0
ざまあ	81.0	36.0	21.0	31.0	26.0	103.0	85.0	52.0	426.0
シブシブ	35.0	33.0	72.0	117.0	142.0	93.0	89.0	27.0	636.0
すげえ	165.0	140.0	34.0	160.0	34.0	26.0	26.0	123.0	708.0
すこ	146.0	124.0	33.0	151.0	29.0	31.0	30.0	122.0	686.0

図 9 単語と感情ごとの合計値

その単語の感情別の合計値をその単語ごとに取得した数値の合計で割り単語の感情別に確率を算出した。その結果の一部を表 1 に示す

6. 結果と考察

実験の結果、単語の感情ごとの確率を算出することが出来た。また各単語の感情の確率でその単語に主に含まれている感情が分かり、一つだけの感情から構成されていないことが分かった。これは「w」や「草」のような笑いの表現も嘲笑や苦笑、爆笑など様々な感情が含まれているからだと考えられる。

6.1 今後の展望

本研究では Cpex を用いた既存の切り抜き動画との比較実験を行い、どれだけシステムで作る動画が内容を短縮した切り抜き動画として見やすいものができるかを調査するものである。その実験を行う過程で必要となる単語の感情

要素を確率として詳細に出すことが出来たので、これを用いた Cpex システムを完成させる。またコメントデータだけでは精密なシーン測定は難しい可能性もあるので、コメントデータではなく動画全体のシナリオ測定や音声の測定なども用いて、よりユーザビリティの高いシステムの完成を目指したいと考える。

謝辞 —

本研究の一部は、JSPS KAKENHI Grant Number JP19K20750 の支援を受けた—

参考文献

- [1] 2021 年版 「YouTube、16 の統計データ」入手先 ([//www.infocubic.co.jp/blog/archives/15518/](http://www.infocubic.co.jp/blog/archives/15518/)).
- [2] 2023 年のショート動画はどうか？最新トレンド調査入手先 (<https://prtimes.jp/main/html/rd/p/000000193.000021986.html>).
- [3] 【時短術あり】動画編集にかかる時間は？ 31 人のアンケート結果を公開！入手先 (<https://fukugyo-free.com/video-editing-time/>).
- [4] 末永 智彦, 早川 智一, 疋田 輝雄 “視聴者の時刻同期コメントを用いた動画の特徴シーンの推定” 第 16 回日本データベース学会年次大会, DEIM Forum 2018 E3-4.
- [5] 鳥海 不二夫, 榊 剛史, 吉田 光男 “ソーシャルメディアを用いた新型コロナ禍における感情変化の分析” 人工知能学会論文誌/35 巻 (2020) 4 号入手先 (<https://www.jstage.jst.go.jp/article/tjsai/35/4/35F-K45/pdf>).
- [6] 【2023 年 3 月更新】YouTube 利用状況データ入手先 (<https://videocloud.jp/topics/7869/>)
- [7] . 2023 年のショート動画はどうか？最新トレンド調査入手先 (linersearchplatform.blog.jp/archives/39951598.html).
- [8] 初頭効果とは？アンカリング・親近効果・ハロー効果との違い、具体例と注意点を解説入手先 (<https://www.sprocket.bz/blog/20220829primacyeffect.html>).
- [9] 尾崎遼太郎† 佐々木史織‡ 清木康† “動画とライブ配信のコメントデータを対象としたリアルタイム時系列感性変化計量・可視化・検索システム” DEIM Forum 2019 C4-1 入手先 (<https://db-event.jpn.org/deim2019/post/papers/90.pdf>).
- [10] R. Plutchik, "The Nature of Emotions", Am. Scientist, vol. 89, pp. 344, 2001.