

F-014

盛り上がり検出のための音声解析の一考察

A Study of Audio Analysis for Hot Spots Detection

赤木 信也*

Shinya Akagi

*NTTデータ先端技術株式会社, Email: akagis@intellilink.co.jp

はじめに

動画の盛り上がり箇所の検出は、ハイライト動画の作成に応用することができ、動画編集を手助けできる技術である。盛り上がりと言で述べても定義は様々であり、「草」や「w」といったコメントに注目したもの[1][2]、笑い声や相槌に注目したもの[3]が存在する。また、映像自動要約の研究として、スポーツなどを対象にしたものがいくつか研究されている[4]。本研究では、YouTubeにアップロードされている切り抜き動画と元動画を利用し、切り抜き箇所が盛り上がり箇所であると仮定して、元動画から切り抜き箇所を検出できないか、音声解析を実施した結果について報告する。

研究の技術領域

本研究の目的は、切り抜き動画やハイライト動画の作成を支援または自動的に実施することにある。本研究の技術領域は下記の通りいくつか分かれており、盛り上がり箇所の検出は1番目の技術領域である。

- (1) 盛り上がり箇所の検出
- (2) 場面の切り分け・結合
- (3) タイトル・説明文の作成

過去研究

YouTubeの動画を対象とした事例としては、tf0101によるhighlight-ranking[2]がある。highlight-rankingでは、チャット上に「草」や「w」が現れる数を10秒間毎に計算し、「草」や「w」が多かった箇所を盛り上がり箇所であると仮定して、前後1分間の動画を上位10件のランキング形式で抽出している。チャット上のコメントを利用する手法のため、該当キーワードのコメントが少ない動画やコメント機能自体がオフになっている動画には適用できないという制約がある。

実験方法

Youtubeの切り抜き動画について、多様なジャンルかつ再生数が多い動画を11件抽出した。再生数が多いということは、切り抜き動画および盛り上がり箇所として適切であると考えられる。動画の内訳は、一人で雑談を行う動画(動画6・動画10)、一人でアクションゲームを行う動画(動画3・動画9)、一人でアクションゲームと雑談を行う動画(動画1)、一人でカードゲームのオンライン対戦を行う動画(動画2)、一人で演奏を行う動画(動画7)、二人で雑談を行う動画(動画5・動画8)、二人でアクションゲームと雑談を行う動画(動画4)、プロ野球の練習試合(実況付き)を行う動画(動画11)である。

切り抜き動画の切り抜き箇所が元動画の盛り上がり箇所の正解であると捉え、元動画の音声波形を取得し、音声波形の振幅の大きい箇所および振幅の差が大きい箇所を用いて、切り抜き箇所の検出ができないか分析した。更に、highlight-rankingを用いて、切り抜き箇所の検出ができないか分析を実施した。

プロ野球の練習試合の動画については、切り抜き動画が選手の成績をピックアップしたものであり、試合の勝敗をピックアップする通常の試合と異なっていたため、正解を変更し、試合の結果および盛り上がりに影響すると考えられるホームラン、盗塁、ピンチ/チャンス時の得点場面/アウトを正解とした。

音声解析手法

まず、librosaという音声処理ツールを用いて、元動画の音声波形を取得する。次に、振幅の大きい箇所および振幅の差が大きい箇所を抽出するため、振幅および振幅の差の大きさに基づくソートを行う。最後に、ソートした結果について、上位1000件をK-mean法を用いて10件の似通った時間軸のクラスタにまとめ、各クラスタの時間軸から元動画の盛り上がり箇所を推定する。動画としては推定箇所から後ろ4分間を抽出し、切り抜き動画と比較を行う。highlight-rankingの事例では、推定箇所の前後1分間を抽出する方法が採用されているが、本研究では推定箇所を分かりやすくするために推定箇所から後ろ4分間の動画を抽出する。動画抽出時間としては、切り抜き動画の平均時間が233秒だったため、4分としている。

実験結果

YouTubeの切り抜き動画計11件(平均動画時間2時間30分)について、元動画の音声の振幅または振幅の差が大きい箇所上位10件の推定結果に含まれるかどうか、および、highlight-rankingの上位10件の推定結果に含まれるかどうかを分析した。

	amplitude									
	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
動画1 7:35-7:39:00	5:00:20 5:04:20	1:43:53 1:47:53	4:32:53 4:36:53	4:08:00 8	3:51:56 3:55:56	4:52:49 23	2:11:42 2:15:42	1:26:11 1:30:11	3:41:54 3:45:54	3:04:38 3:08:38
動画2 1:43:05~1:52:44	7:26:11 26	1:12:04 1:16:04	19:22:23 22	1:51:47 1:55:47	1:35:58 1:39:58	2:09:06 2:13:06	37:19:41 19	1:04:28 1:08:28	1:21:29 1:25:29	45:58:49 58
動画3 28:20~29:30	57:32:1 01:32	48:35:52 35	53:40:57 40	23:16:27 16	41:07:45 07	13:06:17 06	38:25:42 25	36:4:36 4	18:29:22 29	29:14:33 14
動画4 23:47~29:25	10:11:14 11	56:23:1 02:23	1:25:56 1:29:56	1:06:31 1:10:31	2:19:58 2:23:58	2:42:20 2:46:20	1:53:24 1:57:24	37:05:41 05	2:58:37 3:02:37	1:45:34 1:49:34
動画5 09:54~12:53, 26:20~27:01, 32:30~34:55	11:29:15 29	8:41:12 41	16:43:20 43	34:04:38 04	20:41:24 41	57:42:1 42	42:43:47 43	23:31:26 31	31:06:35 06	46:49:50 49
動画6 20:15~26:35	55:11:59 11	43:49:47 49	52:10:56 10	8:44:12 44	22:48:26 48	39:33:43 33	17:51:21 51	34:37:38 37	30:57:34 57	1:02:57 02
動画7 17:10~20:16	58:47:1 02:47	1:19:20 1:23:20	1:13:54 1:17:54	29:19:33 19	1:10:54 1:14:54	5:00:54 05	19:47:23 47	1:22:41 1:26:41	39:36:43 36	1:11:15:15 15
動画8 23:47~37:18	1:00:58 1:04:58	21:34:25 34	5:17:41 17	15:04:19 04	54:25:58 25	27:43:31 43	17:00:21 00	35:50:39 50	10:09:14 09	44:31:48 31
動画9 22:58~24:02	58:08:1 02:08	1:34:41 1:38:41	33:46:37 46	46:08:50 08	9:02:13 02	36:57:40 57	1:03:02 1:07:02	48:26:52 26	53:59:57 59	1:18:54 1:22:54
動画10 1:02:44~1:14:16	56:05:1 00:05	39:59:43 59	1:09:46 1:13:46	23:12:27 12	3:04:01 3:08:01	1:34:01 1:38:01	1:16:27 1:20:27	2:23:03 2:27:03	2:53:52 2:57:52	2:01:18 2:05:18
動画11 38:44, 41:53, 47:52, 53:09, 1:21:38, 2:31:29, 2:34:54, 2:42:21, 2:43:37, 2:47:18, 2:48:49, 2:57:23, 2:59:24, 3:12:53, 3:14:41, 3:16:53	10:37:14 37	1:42:34 1:46:34	16:30:20 30	1:06:28 1:10:28	46:30:50 30	37:28:41 28	2:54:07 2:58:07	3:13:26 3:17:26	3:16:53 3:20:53	1:28:20 1:32:20

		diff amplitude									
		1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
動画1	7:35:38~7:39:00	14353-5038	50738-33142	21142-43253	452349-43653	23	24953-40880	8	64836-65236	11313-11713	
動画2	1:43:05~1:52:44	15739-20324	102014-310835	21213-12140	150550-46464	10543-12918	13918-10943				
動画3	28:20~29:30	5538-59	141018-410645	48342-382642	231627-57311	0131-18322	10215-10615	291433-14			
動画4	23:47~29:25	14329-451149	11846-550059	24358-24758	130417-283232	0432-252129	30070-18432	320136-393943			
動画5	09:54~12:53, 26:20~27:01, 32:30~34:55	46230-56221	112915-29	84112-4132	283232-21	0	521066-105614	10257-292933			
動画6	20:15~26:35	253629-551159	434947-143518	342238-22	41881-10	56	105614-105614	10257-292933			
動画7	17:10~20:16	11708-11946	11124-12144	11344-10235	200304-58001	110715-402844					
動画8	29:46~37:18	17231-241328	542558-402644	140518-210625	273731-353639	475851-65010					
動画9	22:58~24:02	56301-0030	144018-13	02	201124-11755	482762-12051	13133-425146				
動画10	1:02:44~1:14:16	56051-0005	11547-11947	10946-11346	232427-24	12952-59	10415-10513	22303-22753			
動画11	38:44, 41:53, 47:52, 53:09, 1:21:38, 1:31:29, 2:34:54, 2:42:21, 2:43:37, 2:47:18, 2:48:49, 2:57:23, 2:59:24, 3:12:53, 3:14:41, 3:16:53	14631-14631	21917-20503	10551-13218	24721-30011	52	32508-01				

		highlight-ranking									
		1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
動画1	7:35:38~7:39:00	43420-43525	506-11334	640-44356	14418-14804	15034-24858	31849-31954				
動画2	1:43:05~1:52:44	981-981	15157-5242	51335-559	137511-13856	4528-12444	13618-20944	21949-21949			
動画3	28:20~29:30	11561-503	29393-044	57355-840	17471-852	33063-411	25572-702	52575-402	02060-313	06220-727	10441-149
動画4	23:47~29:25	07510-856	13481-453	11358-11503	05440-649	36093-714	40444-149	25047-25152	11115-11220	15952-20057	18471-952
動画5	09:54~12:53, 26:20~27:01, 32:30~34:55	11021-2107	33093-414	19592-104	03000-405	08400-845	53185-423	48124-917	05210-626	28132-918	33433-648
動画6	20:15~26:35	22552-400	42394-344	53265-431	25392-640	11951-210	51185-223	05411-046	33563-501	38483-953	55365-641
動画7	17:10~20:16	10557-10702	50015-106								
動画8	29:46~37:18	25532-658	16091-714	46334-738	14121-1917	32213-326	10051-10156	16411-746	37413-846	11081-213	10300-10405
動画9	22:58~24:02	20342-139	54265-531	42334-338	19342-039	23032-405	29003-005	12111-12216	12323-12428	12939-13044	21102-215
動画10	1:02:44~1:14:16	10217-10322	30738-30838	44344-539	113325-11437	15813-15918	14171-522	37023-807	10904-11009	10305-11114	23922-23127
動画11	38:44, 41:53, 47:52, 53:09, 1:21:38, 1:31:29, 2:34:54, 2:42:21, 2:43:37, 2:47:18, 2:48:49, 2:57:23, 2:59:24, 3:12:53, 3:14:41, 3:16:53										

各動画に着目すると、動画1と動画4は、アクションゲームと雑談を含む動画であり、切り抜き箇所が雑談部分だったこともあってか、いずれの方法でも抽出できていなかった。動画3は、切り抜き箇所が驚きを含む動画であり、音声波形を用いた方法では抽出できていたが、コメントを用いた方法では抽出できていなかった。動画7は、切り抜き箇所が喜びを含む動画であり、音声波形を用いた方法では抽出できていたが、コメントを用いた方法では抽出自体に失敗していた。動画9は、切り抜き箇所が失敗場面を含む動画であり、音声波形を用いた方法では抽出できていなかったが、コメントを用いた方法では抽出できていた。動画10は、切り抜き箇所が悲鳴のような驚きを含む動画であり、いずれの方法でも部分的な抽出には成功していたものの、悲鳴のような驚きを含む箇所の抽出には失敗していた。動画11は、コメント機能が無効のため、コメントを用いた方法では抽出自体ができなかった。振幅を用いた方法において、雑音を抽出してしまったり、ホームランや得点場面などの抽出に失敗していたりしており、部分的な抽出が成功しているように見えるが、実際は正解箇所を多く設定していたことで偶然正解箇所を抽出できていただけだと考えられる。

全体については、正解箇所を多く設定していた動画11を除く計10件の動画について、振幅を用いた方法、振幅の差を用いた方法では10件中7件、highlight-rankingの方法では10件中6件で切り抜き箇所の抽出に成功しており、いずれの方法でも6~7割程度の検出性能があることが示された。

今後について

ゲーム動画(ボス戦,ガチャ,雑談,失敗場面など),映画ドラマ(ホラー,恋愛,コメディなど),演奏動画(独奏,合奏),スポーツ中継動画(野球,サッカー,陸上など)のような多様な分野の動画において,どのような箇所を抽出しやすいか分析することが挙げられる。コメントを用いた盛り上がり検出については,「草」や「w」だけでなく「👏」や「ナイス」,「おお(感嘆表現)」,「おめでとう」といった別の単語を用いた抽出も試す価値があると考えられる。場面の切り分け方法については,盛り上がり推定箇所から前後何秒間を抽出するか調整したり,予め動画を場面毎に分けて盛り上がり箇所が含まれる場面を抽出したりする工夫が必要だと考える。

今回は汎用的な手法を検討したが,精度が高いとは言えない結果となったため,分野最適化の手法を検討しても良いと考える。

まとめ

YouTubeの切り抜き動画と元動画について,音声波形の振幅やコメントを用いて切り抜き箇所の検出ができなにか分析したところ,プロ野球の練習試合動画を除く計10件の動画について,振幅を用いた方法,振幅の差を用いた方法では10件中7件,highlight-rankingの方法では10件中6件で切り抜き箇所の抽出に成功しており,いずれの方法でも6~7割程度の検出性能があることが示された。振幅や振幅の差を用いた方法では,驚きや喜び,拍手音などが抽出されており,それらが切り抜き箇所の抽出にマッチしたものと考えられる。ただし,コメントを用いたhighlight-rankingの方法でしか切り抜き箇所を抽出できなかった動画もあったことから,用途に応じて併用することが必要だと考える。

プロ野球の練習試合動画(動画11)については,ホームランや得点場面などの抽出に失敗していたことから,実況動画と音声の振幅を用いたプロ野球動画の盛り上がり検出は難しいという結果が得られた。

参考文献

- [1] tf0101: YouTube Liveの生放送から盛り上がった箇所を自動抽出するCLIを作った話, <https://qiita.com/tf0101/items/efb2484a0b5b1cdc8291>, 2019-07-08
- [2] highlight-ranking, <https://www.highlight-ranking.com/>
- [3] 河原達也, 須見康平, 緒方淳, 後藤真孝: 音声会話コンテンツにおける聴衆の反応に基づく音響イベントとホットスポットの検出, 情報処理学会論文誌 52 (12), 3363-3373, 2011-12-15
- [4] 望月貴裕: 映像自動要約技術の最新動向, NHK技研R&D2020年夏号, 4-15, 2020-08-15