

# ユーザ投稿型動画サイトのコメントに基づく動画分類手法

11571081 中西 廉(灘本研究室)

あらまし:本研究では,ユーザ投稿型サイトの動画分類手法としてユーザ投稿型動画に付随するコメントを用いた分類手法を提案する.具体的には,ユーザ投稿型動画サイト「YouTube」のゲームのテーマにおいて,「実況」「解説」「RTA」の3つに分類する手法の提案を行う.

## 1. はじめに

近年,YouTube<sup>1</sup>やニコニコ動画<sup>2</sup>といったユーザ投稿型の動画サイトが普及している.それに伴い投稿動画及び動画投稿者が増加することが考えられる.そのため,ユーザが好みの動画を見つけることが難しくなっている.本研究ではYouTubeを対象とし,テーマはゲームとする.ユーザは動画の内容でその動画が好みか否か判断するが,以下に示す動画へのアクセス経路は,タイトルやタグのみで検索されており,動画の内容は考慮されていない.

- YouTube 検索からのアクセス
- Google 検索からのアクセス
- 関連動画からのアクセス
- YouTube トップページからのアクセス
- シェアによるアクセス

内容を考慮して動画を分類することで,よりユーザが求める動画を推薦,検索しやすくなる.本研究では,「マリオカートの解説」の「解説」や「星のカービィの実況」の「実況」など,その動画の主題に対して何を行っているかという部分をジャンルと呼ぶ.

動画のコメントを活用した研究は数多く行われているが<sup>[1][2]</sup>,本研究では動画の内容を考慮するためにコメントを活用することで,ゲームの動画のジャンルを「実況」「解説」「RTA(Real Time Attack)」の3つに分類する手法を提案する.

以下に,提案システムの流れを示す.

- ① ゲームタイトルを入力する
- ② 検索結果の動画群から RTA 動画を抽出する
- ③ 残りの動画群から解説・実況動画を抽出する
- ④ 解説動画と実況動画を分類する

## 2. RTA 動画の抽出

本研究では,RTAは時間を競うことにより,時間に関する特徴語などがあると考え,ルールベースを用いて抽出する.

### 2. 1. RTA 用語の決定

手動で分類した「実況」「解説」「RTA」「その他」

各 100 件の動画に付随したコメントを分析し,RTAによく用いられている単語を RTA 用語と呼ぶ.

この RTA 用語を用いて RTA 動画を抽出する.そこで RTA 用語を決定するために,取得した動画のコメントから形態素解析器 JUMAN<sup>3</sup>を用いて名詞を取得する.次に「RTA」に含まれる名詞と「実況」「解説」「その他」に含まれる名詞を比較し,「実況」「解説」「その他」と重複した名詞を「RTA」の名詞群から削除する.残った RTA の名詞群から投稿者及びゲームに依存する語は不要と考え,それらに関する固有名詞を削除する.得られた RTA の名詞群の tf 値を求め,上位  $\alpha$  件を RTA 用語とする.本研究では  $\alpha$  を 15 件とした.これだけでは RTA の専門用語を抽出できないと考え,さらに RTA の専門用語集から英語を省き,残った用語も RTA 用語とする.その結果,31 語の RTA 用語を決定した.表 1 に RTA 用語を示す.

表 1. RTA 用語

タイム	ショートカット	シヨトカ	最速
記録	更新	世界記録	WR
wr	短縮	走者	タイマー
ランナー	100%	any%	タイムアタック
TA	バグ	グリッチ	再走
リセット	リセ	ルート	仕様
秒	レギュレーション	レギュ	やりこみ
スピードラン	スピラン	プラットホーム	

## 2. 2. RTA 動画の決定

ある動画に対して,コメントに出現する RTA 用語の出現回数の総和をその動画のコメント総数で割った値がある閾値  $\beta$  以上のものを RTA 動画とする.本研究では  $\beta$  を 0.1 とする.

## 3. 実況動画, 解説動画の抽出

実況・解説の抽出は,実況・解説の特徴語が含まれるかどうかにより決定する.実況・解説の特徴語をここでは実況・解説用語と呼ぶ.

### 3. 1. 実況・解説用語の決定

実況・解説用語の決定手法は RTA 用語の抽出手法と同様である,RTA 用語決定では,「RTA」のコメントに用いられている名詞と「実況」

<sup>1</sup> YouTube: <https://www.youtube.com/>

<sup>2</sup> ニコニコ動画: <https://www.nicovideo.jp/>

<sup>3</sup> JUMAN: <http://nlp.st.i.kyoto-u.ac.jp/>

「解説」「その他」に含まれている名詞を比較したが、ここでは「実況」「解説」に含まれている名詞と「RTA」「その他」に含まれている名詞を比較し重複を削除し実況・解説用語を決定する。決定した実況・解説用語は1,742語である。表2に実況・解説用語の一部を示す。

表2. 実況・解説用語の一部

参考	初心	攻撃	最近	キャラ	攻略
楽チン	手本	要素	尊敬	判定	注意
発想	練習	腹筋	期待	応援	安定

### 3. 2. 実況動画と解説動画の決定

RTA動画を削除した動画群に対し、実況・解説用語の出現回数の総和をその動画のコメント総数で割った値がある閾値 $\gamma$ 以上の動画を実況と解説動画と決定する。本研究では、 $\gamma$ を0.5とした。

### 3. 3. 実況動画と解説動画の分類

実況と解説動画の分類は叫喚メッセージ<sup>[3]</sup>の有無で決定する。実況と解説動画に対し、叫喚メッセージが含まれているものを実況動画、含まれていないものを解説動画とする。ここで、叫喚メッセージとはユーザが強い感情を表す際に用いる母音を連続させる語である。以下の条件を満たすものを叫喚メッセージとする。

- ・語尾の母音が3回以上繰り返して付加されているもの
- ・母音は大文字小文字を区別しない
- ・母音はひらがな、カタカナの大小文字全て

## 4. 評価実験

提案手法の有効性を示す為に、YouTubeの動画分類と提案手法の比較実験を実施する。

### 4. 1. 使用データ

YouTube Data API v3<sup>4</sup>でゲーム名「ダークソウル」「ブラッドボーン」「ゼルダの伝説」をクエリとし、得られた日本の動画300本を用いる。正解データは、被験者3名が手動で「実況」「解説」「RTA」「その他」に分類したデータとする。

### 4. 2. 実験結果と考察

提案手法と既存のYouTube検索の実験結果を比較したものを表3、表4、表5に示す。

表3と表4のRTAにおいて提案手法より既存手法のF値が高くなっている。これは、日本の動画だがタイトルがすべて英語で記載されている動画に付随するコメントの中に英語のみで記載されたコメントが多数あり、RTA用語に含まれる語が少なかったためと考えられる。RTA以外のジャンルにおいてほぼ全ての再現率が既存の手法

より提案手法のほうが高い。これはタイトルにジャンルがわかる言葉が入っていても、コメントからどのジャンルか推定でき、既存手法ではその他に分類されるが、提案手法ではその他以外のジャンルに分類できる為と考えられる。

表3. 「ダークソウル」

手法	ジャンル	適合率	再現率	F値
You Tube 検索	実況	0.91	0.33	0.49
	解説	0.89	0.45	0.60
	RTA	0.60	1.00	0.74
提案手法	実況	0.89	0.59	0.71
	解説	0.61	0.64	0.63
	RTA	0.42	0.33	0.37

表4. 「ブラッドボーン」

手法	ジャンル	適合率	再現率	F値
You Tube 検索	実況	0.79	0.33	0.47
	解説	0.60	0.29	0.39
	RTA	0.66	1.00	0.8
提案手法	実況	0.91	0.39	0.55
	解説	0.73	0.45	0.56
	RTA	0.375	0.75	0.5

表5. 「ゼルダの伝説」

手法	ジャンル	適合率	再現率	F値
You Tube 検索	実況	0.97	0.64	0.77
	解説	0.54	0.42	0.47
	RTA	0.42	0.6	0.5
提案手法	実況	0.94	0.51	0.66
	解説	0.68	0.46	0.55
	RTA	0.66	0.8	0.72

## 5. まとめと今後の課題

本研究では、ゲームの動画を対象とし既存の手法では分類できない動画を、付随するコメントを用いることで分類する手法を示した。具体的には、RTA用語と実況・解説用語を作成しRTA動画の分類及び実況・解説の分類を行った。さらに叫喚フレーズを用いることで、実況と解説の分類を行った。今後の課題はゲームに限らず様々なテーマを考慮する必要がある。

### 参考文献

- [1]青木秀憲, 宮下芳明, "ニコニコ動画における映像要約とさび検出の試み", 情報処理学会研究報告ヒューマンコンピュータインタラクション, pp.37-42, 2008
- [2]中村聡史, 田中克己, "印象に基づく動画検索", 研究報告ヒューマンインタラクション, pp. 77-

<sup>4</sup> YouTube Data API v3:  
<https://developers.google.com/youtube/v3/>

84, 2009

[3] 浅井洋樹, 秋岡明香, 山名早人”きたあああ  
あああああああああああ!!!! 1 1:マ  
イクロブログを用いた教師なし叫喚フレーズ抽  
出”, DEIM Forum, A4-4, 2014.