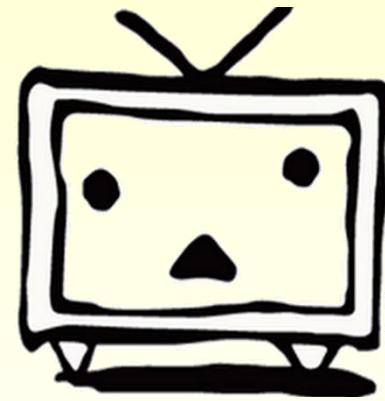


ユーザ投稿型動画サイトの コメントに基づく動画分類手法

灘本研究室 11571081 中西 廉

背景

You **Tube**



ユーザ投稿型動画サイト

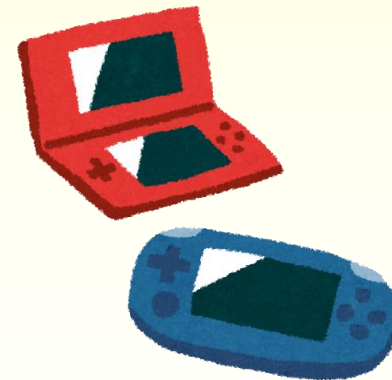
背景

日本国内のユーザーが
普段視聴している
無料動画コンテンツ
サービスランキング *1



You **Tube**

2018年動画の投稿が
公式に認められた *2

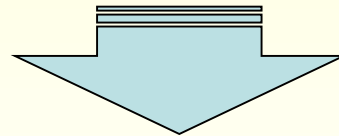
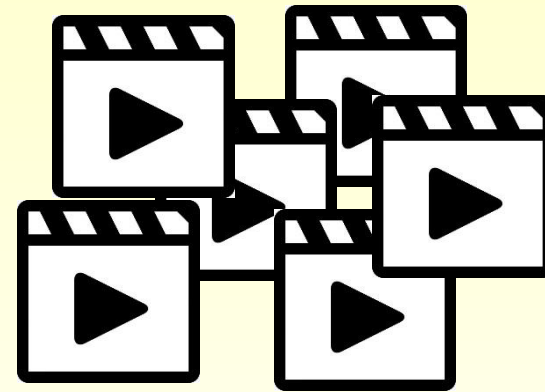


ゲームの動画

*1:Tubular
<https://blog.brkr.jp/youtubeupdate/>

*2:VIP初心者がゲーム実況をするには@WIKI
<https://www18.atwiki.jp/live2ch/pages/157.html>

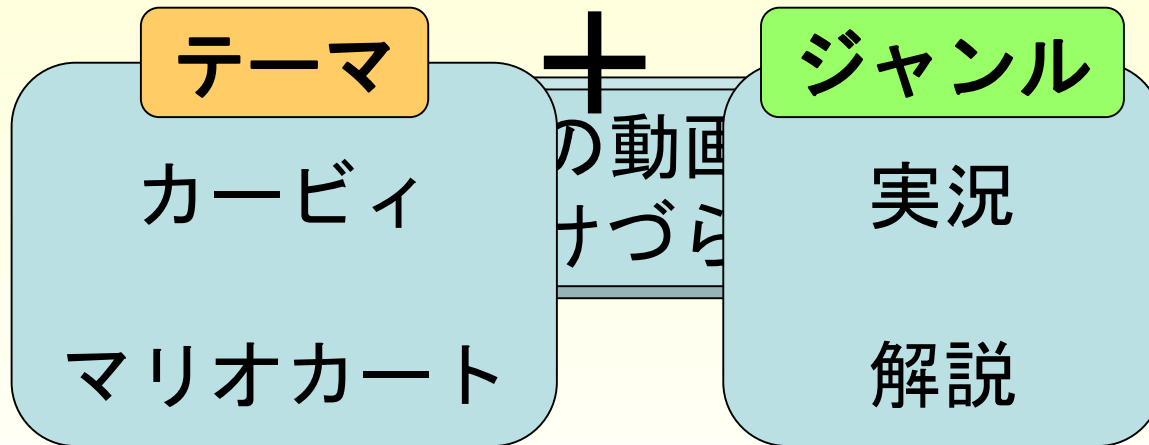
背景



好みの動画を見つげづらい

背景

「カービィ」の「実況」
「マリオカート」の「解説」



テーマ

動画の題材となるもの

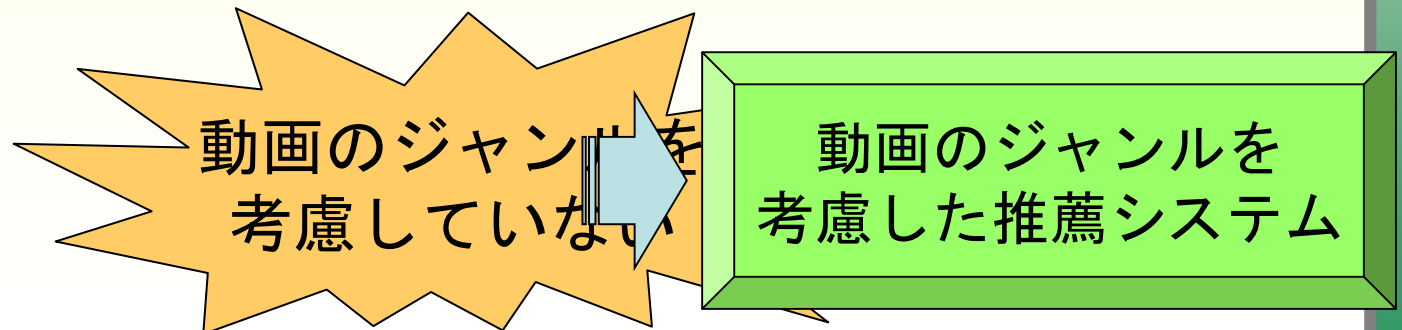
ジャンル

題材に対し何をしているか

背景

YouTubeの動画にアクセスする経路

アクセス経路	アクセスキー
YouTube検索からのアクセス	タイトル
Google検索からのアクセス	説明文
関連動画からのアクセス	タグ
YouTube トップページからのアクセス	チャンネル
シェアによるアクセス	



目的

ジャンルを考慮した
動画分類手法の提案



ジャンルの定義

実況

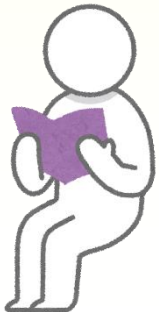
投稿者の言葉や文字があり実際にプレイしている動画

解説

ゲームについての説明をメインにしている動画

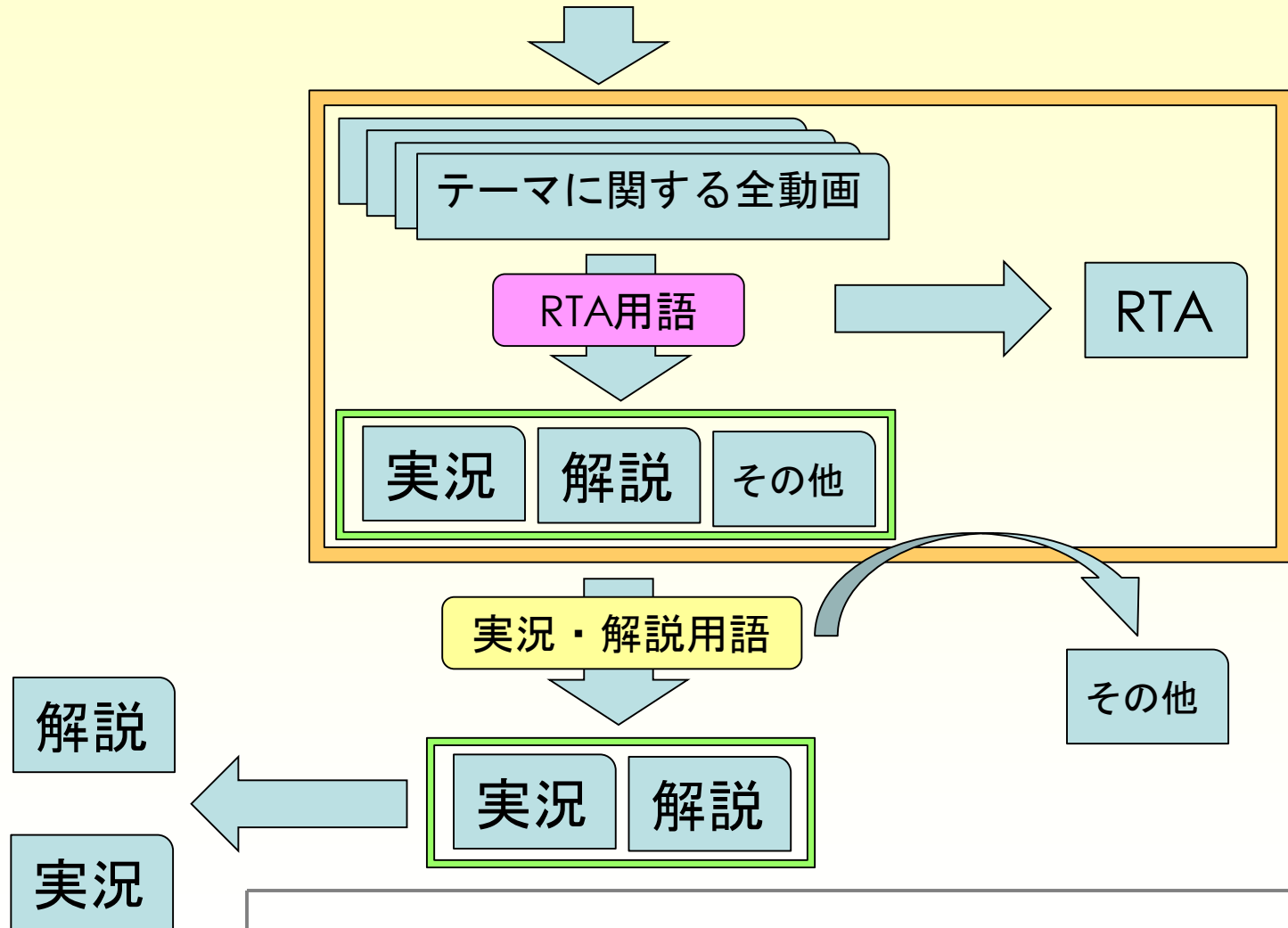
RTA

Real Time Attack
ゲーム開始からクリアまでにかかった現実時間の短さを競う競技



本研究の流れ

テーマを入力



RTA用語の作成

実況 100件
 解説 100件
 RTA 100件
 その他 100件

形態素解析器 JUMAN

名詞を取得



重複している語を削除
 人物に関する固有名詞を削除
 ゲームに関する固有名詞を削除
RTA用語 31語

RTA Tf値の上位15件

+

Speedrun用語集^{*4}

タイム ショートカット ショトカ 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% 再走
 any% タイムアタック TA バグ 秒
 リセット リセ ルート 仕様 グリッチ
 レギュレーション レギュ スピードラン
 スピラン プラットホーム やりこみ

*4:Speedrun用語集

https://ch.nicovideo.jp/gunsou_pl-mi/blomaga/ar1103535から一部引用

RTA動画分類

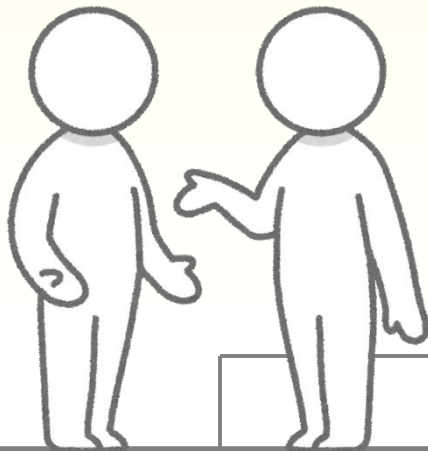
RTA

動画のコメントに出現する
RTA用語出現回数の総和

動画のコメント総数

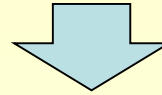
= α

$$\alpha \geq 0.1$$



本研究の流れ

テーマを入力



テーマに関する全動画

RTA用語

RTA

実況

解説

その他

実況・解説用語

その他

解説

実況

実況

解説

実況・解説用語の作成

実況	100件
解説	100件
RTA	100件
その他	100件

形態素解析器 JUMAN*5

名詞を取得

実況

解説

その他

RTA



重複する単語を削除

実況

解説

その他

重複する単語を削除

人物に関する固有名詞を削除

ゲームに関する固有名詞を削除

実況・解説用語

*5:JUMAN:<http://nlp.st.i.kyoto-u.ac.jp/>

実況・解説動画分類

実況・解説用語 例

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

1742語

実況・解説用語

実況

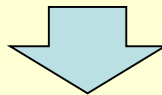
解説

$$\frac{\text{動画のコメントに出現する
実況・解説用語の出現回数の総和}}{\text{その動画のコメント総数}} = \beta$$

$$\beta \geq 0.5$$

本研究の流れ

テーマを入力



テーマに関する全動画

RTA用語



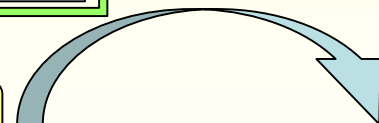
RTA

実況

解説

その他

実況・解説用語



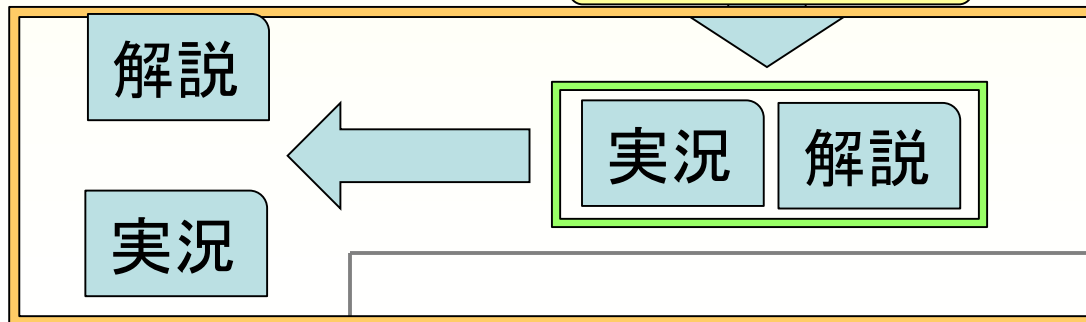
その他

解説

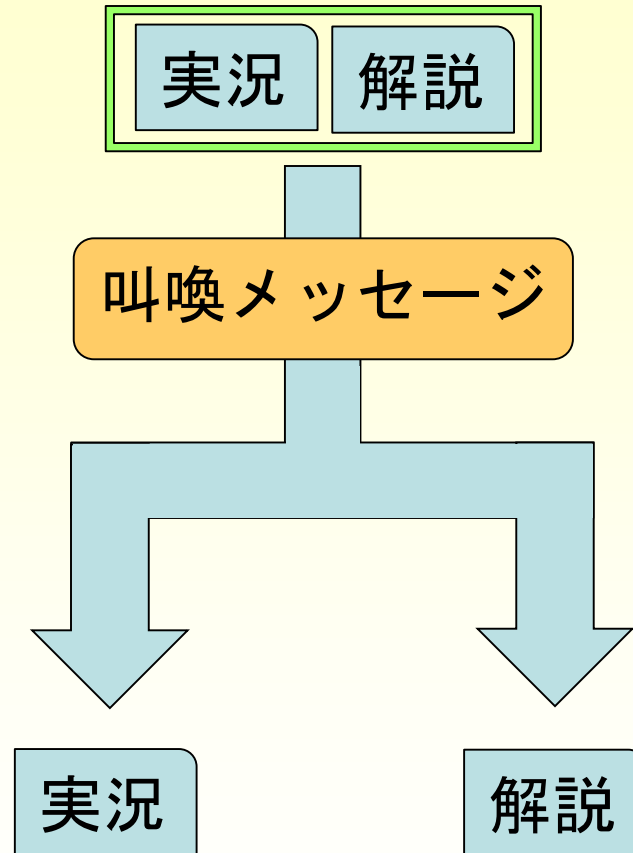
実況

解説

実況



分類の流れ



実況動画と解説動画の分類

叫喚メッセージ

実況 叫喚メッセージを含むコメントがされている動画

解説 叫喚メッセージを含むコメントがされていない動画



評価実験

目的

提案手法の有効性を示す為に、ベースライン (YouTubeの動画分類) と提案手法の比較実験を実施

使用データ

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データ各300件を使用

20代男性の被験者3人に手動で

「実況」「解説」「RTA」「その他」に分類してもらい
正解データとした

結果と考察:ダークソウル

手法	ジャンル	適合率	再現率	F値
ベースライン	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

- RTAではベースラインのほうが精度が高い
→英語のタイトル、コメントが多い
- 実況・解説は提案手法が再現率が高い
→ベースラインではその他に含まれる動画も分類できる

結果と考察:ブラッドボーン

手法	ジャンル	適合率	再現率	F値
ベースライン	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

- RTAではベースラインのほうが精度が高い
→英語のタイトル、コメントが多い
- 実況・解説は提案手法が再現率が高い
→ベースラインではその他に含まれる動画も分類できる

結果と考察:ゼルダの伝説

手法	ジャンル	適合率	再現率	F値
ベースライン	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

- RTA動画も提案手法のほうが精度が高い
→日本語のタイトル、コメントが多い
- ベースラインの実況の再現率が高い
→タイトルに実況とわかる動画が多い

まとめと今後の課題

□ まとめ

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

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

ゲーム以外のテーマへの対応

日本の動画以外への対応