

# 視聴者コメントとリッジ回帰を用いた 動画共有サービスのジャンルごとのニーズ調査

田中 耀祐† 松田 源立†

成蹊大学 理工学部 情報科学科†

## 1. はじめに

近年、オンラインでのコンテンツ視聴が拡大し、動画投稿者の需要が高まっている。しかし、動画投稿者は競争が激しく、再生回数を獲得することは容易ではない。そのため、視聴者の意見や感想が反映されているコメント欄を分析することは再生回数を伸ばす手段として重要である。本研究では YouTube のコメント欄と再生回数をジャンルに分けて収集し、再生回数を高めている名詞を明らかにすることによって、視聴者が動画にどのようなニーズを求めているかを調査することを目的とする。

## 2. 先行研究

[1]ではニコニコ動画の VOCAROID 関連の動画についてコメント数、動画再生数、投稿時期等を用いて、再生回数を推定する線形回帰分析を行っている。[2]では、多言語展開された同一 YouTube コンテンツのコメントを対象とした言語ごとの特徴語抽出方式について示した研究である。[1]及び[2]は一つのジャンルや動画に重点を置いて分析した研究となっている。[3]では動画をジャンルごとに分けてコメントの特徴を分析し、再生回数の多い動画と少ない動画を比較することで各ジャンルの視聴者の特徴を考察している。本研究では[3]を基に同ジャンルに着目し、名詞リストやストップワードを参考にして分析を行った。

## 3. 研究手法

### 3.1. データ収集

YouTube Date API の機能で検索する語句を指定し、「ゲーム実況」、「ドッキリ」、「大食い」、「商品紹介」、「漫才」の 5 つのジャンルについて再生回数が高い動画から順に動画を収集し、各動画からコメント欄と再生回数を取得した。各ジャンルの動画数とコメント数を表 1 に示す。

表 1 各ジャンルの動画数とコメント数

ジャンル	動画数	コメント数
ゲーム実況	403	135002
ドッキリ	325	257250
大食い	330	184696
商品紹介	122	82690
漫才	144	40957

Analysis of Genre-Dependent Effects of Viewer Comments in Video Sharing Site by Ridge Regression

†Yosuke Tanaka, Yoshitatsu Matsuda

Department of Computer and Information Science, Faculty of Science and Technology, Seikei University

### 3.2. 形態素解析

形態素解析エンジンである Janome を使用し、各動画のコメント欄に出現する名詞のみを抽出した。いずれの動画にも出現する名詞や局所的にしか使われない名詞は分析対象から除いた。また、人気であることは自明であるゲーム名や動画投稿者の名前等の固有名詞はストップワードとして目視で名詞リストから除外した。合計で 2982 種類の単語を使用した。

### 3.3. リッジ回帰

リッジ回帰とは、切片を除く回帰係数の 2 乗和をペナルティ項としてとる正則化法の一つであり、線形回帰モデルでの強い相関によって生じる回帰係数の推定値の不安定性を回避する手法である[4]。

本研究では「ゲーム実況」、「ドッキリ」、「大食い」、「商品紹介」、「漫才」、「全ジャンル」の 6 つのジャンルについてリッジ回帰を行った。目的変数は動画の再生回数を  $\log_{10}$  で対数変換した数値とし、説明変数は対象の名詞が動画のコメント欄に含まれている場合 1、含まれていない場合 0 とした。また、リッジ回帰は正則化の強さを制御するハイパーパラメータが必要であるため、5 分割交差検証法で全ジャンルを除く 5 つのジャンルの決定係数の平均値の合計を計算し、合計値が最大となるハイパーパラメータを共通値として用いた。本研究では Python のライブラリである scikit-learn を用いてリッジ回帰を行った。

### 3.4. ケンドールの順位相関係数

リッジ回帰から得られた名詞の偏回帰係数を降順にソートし、偏回帰係数の一番高い名詞から順に順位付けをした。本研究では Python のライブラリである SciPy を用いてケンドールの順位相関係数の計算を行った。

## 4. 結果・考察

### 4.1. ジャンルごとのリッジ回帰の決定係数

3.3 節の共通のハイパーパラメータを用いて、リッジ回帰の決定係数を 5 分割交差検証法で求めた結果を表 2 に示す。

表 2 ジャンルごとのリッジ回帰の決定係数

ジャンル	決定係数
ゲーム実況	0.473
ドッキリ	0.435
大食い	0.443
商品紹介	0.576
漫才	0.547
全ジャンル	0.490

全てのジャンルにおいて決定係数が 0.5 程度であり、単純な線形回帰である程度説明可能であることが示された。

「ゲーム実況」及び「全ジャンル」について 5 分割交差検証法の一部のテストデータにおける予測視聴回数と実際の視聴回数の比較を図 1、2 に示す。図示においても、再生回数の対数がコメント欄の名詞生起の線形関数により推定可能であることが示されている。この結果より、コメント欄の名詞の生起は再生回数の対数を予測する上で有用であることが示唆される。

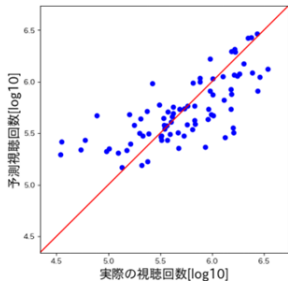


図 1 ゲーム実況

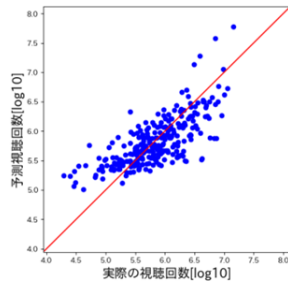


図 2 全ジャンル

#### 4.2. ジャンルごとの偏回帰係数上位の名詞

リッジ回帰の各ジャンルの偏回帰係数上位 10 位の名詞を表 3 に示す。「ゲーム実況」ジャンルについては「オチ」等の名詞が見られた。「オチ」に関しては「オチが分かっている面白」というコメントが多く見受けられた。分かりやすいオチでも視聴者の笑いを誘うことができるため、オチを作ることはゲーム実況動画を制作する上で重要な要素であることが考えられる。「ドッキリ」ジャンルについては「ショート」等の名詞が見られた。「ショート」に関しては「ショートから」というコメントが多く見受けられた。視聴者はドッキリジャンルの動画選択をする中で YouTube Shorts を参考にしていることが考えられる。そのため、本編の動画の見どころなどを切り抜き、short 動画を上げることは再生回数を伸ばす要因となるということが示唆される。「大食い」ジャンルについては「Wow」等の名詞が見られた。

「Wow」は外国人が驚く際に用いられていた。海外の人も楽しめるかつインパクトのある大食いが再生回数を伸ばす上で重要であることが考えられる。「商品紹介」については「髪」等の名詞が見られた。「髪」に関しては髪の悩みについてのコメントが多く見受けられた。髪に対する問題を解決する商品を多くの視聴者が求めているということが示唆される。「漫才」ジャンルについては「テンポ」等の名詞が見られた。「テンポ」に関しては「テンポがいい」というコメントが多く見受けられた。漫才ジャンルでは視聴者を飽きさせないテンポ作りが再生回数を伸ばす上で重要だと考えられる。「全ジャンル」については「ドッキリ」等の名詞が見られた。総合的に「ドッキリ」関連等のコンテンツが視聴者にとって需要があると考えられる。

これらの分析結果を基に、動画制作においてより効果的なコンテンツの提供が可能となり、視聴者に高い満足度を提供できると考えられる。

表 3 各ジャンルの偏回帰係数上位 10 位の名詞

ゲーム実況	ドッキリ	大食い	商品紹介	漫才	全ジャンル
ゲーム	なに	Wow	髪	テンポ	ドッキリ
チャンネル登録	女	♥	家族	天才	ゲーム
男	さそう	日本人	離婚	くだり	♥
曲	髪型	w	奴	二人	テンポ
子供	妹	食費	wwwww	神	お父さん
オチ	体	腹	仲	笑い	なに
好きw	ショート	バカ	デザイン	www	天才
解説	周り	カット	天才	wwwww	二人
♥	お父さん	あなた	デカ	本物	本物
彼女	好きw	キムチ	髪型	敵	ツッコミ

#### 4.3. 2 ジャンル間のケンドールの順位相関係数

2 ジャンル間のケンドールの順位相関係数を表 4 に示す。「全ジャンル」以外の異なるジャンル間の相関係数は 0 に近く、ジャンルによってコメントにおける重要名詞には大きな差異があることが示された。

このことから動画作成時における適切なジャンル設定の重要性が強く示唆される。

表 4 2 ジャンル間の順位相関係数

	ゲーム実況	ドッキリ	大食い	商品紹介	漫才	全ジャンル
ゲーム実況		0.055	0.002	0.052	0.047	0.339
ドッキリ	0.055		0.024	0.061	0.047	0.273
大食い	0.002	0.024		0.031	0.014	0.210
商品紹介	0.052	0.061	0.031		0.061	0.220
漫才	0.047	0.047	0.014	0.061		0.222
全ジャンル	0.339	0.273	0.210	0.220	0.222	

#### 5. まとめ

本研究では動画共有サービスのコメント欄と再生回数についてジャンルを分けて収集し、各ジャンルにおいてリッジ回帰を用いてコメント欄の名詞の生起から再生回数の対数が推定可能であることを示した。更に、リッジ回帰により各ジャンルでの再生回数に影響を与える重要名詞を抽出し、視聴者が動画に求めているニーズを発見することができた。また、各ジャンル間での重要名詞の共通性を順位相関係数を用いて調査し、ジャンルによってコメントにおける重要名詞には大きな差異があることを示した。再生回数はコメントだけでなくサムネイルや投稿時間等の影響も受けるため、今後はそれらのデータを追加することを検討している。

本研究は JSPS 科研費 JP21K12036 の助成を受けたものである。

#### 参考文献

- [1] 柴田 知親, 伊東 栄典, “回帰分析による CGM 動画再生回数推定”, 九州大学学術情報リポジトリ (2018)
- [2] 西岡 実希, 岡田 龍太郎, 峰松 彩子, 中西 崇文“多言語展開された同一 YouTube コンテンツのコメントを対象とした言語ごとの特徴語抽出方式”, 第 84 回情報処理学会全国大会(2023)
- [3] 手塚 颯太, “ジャンルと再生回数に基づく YouTube コメントの特徴分析”, 成蹊大学理工学部情報科学科 卒業論文(2022)
- [4] 小西 貞則, 多変量解析入門—線形から非線形へ. 岩波書店 (2010)