

動画共有サイトにおけるコメント語彙の時系列解析手法の提案

柳田 雄輝[†] 若林 啓[†] 佐藤 哲司[†]

[†] 筑波大学 〒305-8550 茨城県つくば市春日 1-2

E-mail: †s1711576@s.tsukuba.ac.jp, ††kwakaba@slis.tsukuba.ac.jp, †††satoh@ce.slis.tsukuba.ac.jp

あらまし 動画共有サイトでは、動画のタイムラインに沿ってコメントを投稿する形態がある。本研究では、コメントを動画の内容に依存せず出現する表現とそうでない表現に分け、それらの語彙と投稿日時の時系列分布の関連性について分布の傾向をもとに分析したので報告する。

キーワード 動画共有サイト, コメント解析, 語彙の分布

1 はじめに

インターネット上には、Youtube¹やニコニコ動画²など、様々な動画共有サイトが存在し、自作の動画を投稿できだけでなく、既存の動画にコメントを付けることもできる。中でも、図1に示すニコニコ動画[1]の視聴画面例³では、動画のタイムラインに沿ってビデオ画面上にコメントが重畳表示されている。この形態で視聴者は、ビデオの再生と同期してコメントを閲覧し、コメントによって動画内容の補足情報を得る、あるいは、画面上に流れるコメントの多寡から盛り上がり箇所を把握することができる。一見すると意味が無い「www」のような文字列も、笑いを表す記号として視聴者コミュニティの中で共通の認識が生まれ、その認識を共有するという二次著作としての側面もコメントにはあると言える。一方で、動画の内容とは関係がないと考えられるコメントが多数投稿され画面を埋め尽くすことで、コメントが動画視聴の妨げとなってしまう側面もある。



図1 ニコニコ動画の視聴画面例

本研究では、ニコニコ動画に投稿されたコメントを対象に、コメントに出現する語彙とその時系列分布を、高頻度語のリスト化と正規表現によって解析する。これによって語彙の時系列分布から視聴者のコメント投稿行動を解明するとともに、動画

の内容に依存しないコメントを推定し、新たなフィルタリング機能の作成に役立てられると期待できる。

2 関連研究

ニコニコ動画のコメントを取り扱った研究は広く行われている。前島らはコメントによってサムネイル動画の自動生成を試みている[2]。そこでは、サムネイル動画の印象評価がうまくいかなかった原因の一つとして、その動画内に「www」という語が頻出するため、本来抽出されるべき語を抽出できなかったことを理由に挙げている。本研究は、このような頻出語が分析を困難にしている研究への知見提供の役割も期待される。平澤らはノベルティの高い動画を発見、推薦するために、コメントを形態素解析して得られた語のうち高いTF-IDF値が得られた語と「もっと評価されるべき」タグ、再生数に対するコメントやマイリストの割合に着目する手法を提案している[3]。彼らはこの研究の限界としてタグの多様さを挙げており、ニコニコ動画の分析におけるテキストデータの重要性を示唆している。

動画のタイムラインに沿ってコメントできる形態に着目した研究もある。例えば、齋藤らはタイムライン上での広告動画の挿入タイミングについて、単位時間を1秒としてコメント数の分散から広告挿入箇所を決定する手法を提案している[4]。佃らの研究[5]では、タイムライン上に登場人物の活躍シーンを定義している。これによって、活躍シーンに沿って付与された人物名を含むコメントと、視聴者の動画に対する印象情報の2つを用いた動画検索の推薦システムを実装している。このシステムでは人物名のような動画あるいはカテゴリ依存の語に注目しており、本研究のようにカテゴリに依存しない高頻度語に焦点を当てたものではない。山本らは楽曲動画のサビ部分に付与されたコメントがその楽曲の印象を表すと仮定し、楽曲動画の印象推定を行っている[6]。ここでは、コメントを形態素解析して形容詞と形容動詞を抽出した上で、繰り返し文字の正規化を行っており、本研究と手法が類似している。しかし、山本らの研究はVOCALOID楽曲のみを対象とした分析であり、カテゴリを跨いだ分析は行っていない点、TF-IDFによって動画の内容をよく表現するコメントに重み付けをしている点で異なる。

ダイジェスト動画の作成を目的とした磯貝らの研究では、コ

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

2 : <https://www.nicovideo.jp/>

3 : <https://www.nicovideo.jp/watch/sm36230032>

メントが多いにも関わらず不要であるシーンを取り除くため、「w」を含むコメントを抽出することで重要度の低いシーンを推定する手法が提案されている[7].「w」を「面白いシーンに対するリアクションとして笑いの意味を込めた」語である[7]と説明しているが、本研究では「w」に限定せず、出現頻度が高く動画の内容に依存しないコメントを推定する分析を行う。ここで挙げた研究はいずれも動画のタイムラインとしての時系列に着目しているが、本研究ではタイムラインとしての時系列に加えて、コメントが投稿された日時における時系列にも着目して分析する。

3 コメント語彙の時系列解析手法

3.1 分析の流れ

ニコニコ動画では日常で使われている文章とは異なる文がコメントとして投稿されている。そのため、コメントの出現語を見るだけでは語の意味を読み取れないこと、形態素解析のみでは類似する表現を拾いきれないことの2点が分析を行う上で課題となる。本研究では、1点目の課題については、各動画について頻度の高い語をリスト化することで、各コメントが動画の内容に依存するものかどうかを判定する手法を提案する。これによって各語が動画の内容に対するコメントなのか、動画に対する「賑やかし」としてのコメントなのかを区別する。2点目の課題については、コメントについて形態素解析を行った後、正規表現を用いてサイト特有の表現を整形することで類似表現をまとめられるようにする。

3.2 分析に用いるデータ

国立情報学研究所情報学研究データリポジトリにて公開されているニコニコ動画コメント等データを用いて分析を行う[8].このデータセットでは、ニコニコ動画のサービス開始日である2007年3月6日から2018年11月8日までに付与された約38億のコメントが公開されており、表1のように、各動画について付与されたコメント投稿時間、コメント本文、書き込み再生位置、コマンドがJSONファイルで提供されている。書き込み再生位置はそのコメントが投稿された再生時刻、すなわち動画の再生開始からの経過秒数を示す値で、0.01秒を単位時間としてデータセットに記録されている。コマンドとは投稿するコメントのサイズや色などを投稿者が設定できる機能のことである。大半のコメントはコマンドを指定せずに投稿されるため、本分析ではコマンドについてはユーザのコメント投稿行動にはほとんど影響を与えないと仮定する。

サイトに投稿されてから十分に時間が経過しているコメントも分析対象とするため、ここでは2013年から2015年までに投稿され、総コメント数が10,000以上の動画に付与されたコメントから抽出して分析する。なお、カテゴリに依存せず投稿されるコメントを推定するため、2013年から2015年当時において代表的といえるカテゴリを複数選択した上でコメントを抽出する。

3.3 分析する語の抽出

ニコニコ動画の代表的なカテゴリであるゲーム、アニメ、歌ってみた、科学の4つのカテゴリに含まれる動画を5本ずつ無作為に抽出する。これは、動画のカテゴリによって付与されるコメントに異なる傾向が見られるであろうとの考えから、内容的に大きく異なるカテゴリを選択することとした。抽出した各動画に付与されているコメントをMecab⁴を用いて形態素解析し、非自立語を除く名詞、形容詞、接尾語と非自立語を除く動詞を原形にして抽出、各語の出現頻度とタイムライン上での投稿位置、投稿日時を集計する。辞書にはmecab-ipadic-neologd⁵を用いることで流行り言葉やサイト特有の表現に対応できるようにした。

形態素の抽出が完了したら、各動画について形態素を出現頻度の降順に並べ、文字の繰り返しが見られる上位語があれば正規表現を用いて整形する。例えば「www」のような同一の文字や記号が2回以上繰り返している場合は、繰り返しが2文字になるように「w」を置き換えて累計する。繰り返される文字がアルファベットや数字、記号などの場合は、全角と半角を同一のものとみなすことにする。

上記の方法で各語を再度累計し、動画やカテゴリに依存せず高頻度で出現する語を推定する。ここで推定したカテゴリ非依存の語については、コメント付与時間における時系列分析と動画のタイムラインにおける時系列分析を行う。コメント付与時間については、動画が投稿されてからコメントが付与されるまでの日時とDF値やカテゴリとの関係を明らかにし、動画のタイムラインについては、タイムライン上におけるコメント付与の過程から視聴者のコメント投稿行動を分析する。

3.4 コメント付与時間における時系列分析

コメント付与時間の時系列分析では、各動画内において出現頻度(TF値)の高い語に文書頻度(DF値)を加算していくことで、DF値の大きさを動画に依存せず登場しやすい語彙であるかどうかの指標とする。まず、先の方法で累計した語について各動画でのTF値が高い上位10語をその動画で頻出する語と考え、DF値を1加算する。その後このカウントを4カテゴリ5本ずつ計20動画に対して行い、各コメントのDF値を算出する。この際、DF値がどのカテゴリによって加算されたものなのかを保持しておくことで、コメント投稿日時と語彙の関係性にカテゴリも加味して分析できるようにした。

次に、DF値が1以上となった語についてヒートマップを作成することで時系列分布を示す。ヒートマップには縦軸にDF値を順に並べた各語、横軸に動画が投稿されてからどの程度経ったかを示す日時区分を表し、語ごとにどの日時区分に頻出しているのかを明らかにする。日時区分については、動画が投稿されてから1日(24時間)以内、1日以上1週間以内、1週間以上1ヶ月以内、それ以降の4区分で区切ることとする。出現した各コメントについて、そのコメントがどの日時区分に投

4: <https://taku910.github.io/mecab/>

5: <https://github.com/neologd/mecab-ipadic-neologd>

表 1 コメントデータの仕様 (株式会社ドワンゴ (2016) のデータ仕様書より)

キー	値	値の型	例
date	コメント投稿時間 (UNIX 時間・秒)	Number	1173122885
content	コメント本文	String	"きたwwwwwwwwwwwwwwwwww"
vpos	書き込み再生位置	Number	1061
command	コマンド	String(nullable)	"big green"

稿されたものなのかを計算し、各区分にどの程度その語が出現したのかを求める。なお、動画が投稿されてからの日時を計算するにあたっては動画が投稿されて最初についたコメントと日時の差分をとることで近似して求めた。

各語の TF 値を 4 区分に分類したら、値を正規化して TF 値が 1 としたときの割合を示すようにした。これによって語彙間の比較が可能となる。加えて、各時間区分については標準化を行い、各語の特徴がより明らかになるようにした。以上の手順で、各語が動画が投稿されてからどの程度経ってから付与されたコメントに含まれるものなのかを判定、全 20 動画についてのヒートマップを作成する。これに加え、各カテゴリについても同様の手順でヒートマップを作成する。調査動画は各カテゴリ 5 本ずつとしているため、この場合では DF 値は最大で 5 となる。

4 実験結果

コメントを集計して得られた出現頻度の上位 10 語について、表 2 に全動画での累計結果を、表 3 と表 4 にカテゴリごとの累計結果を示す。なお、Mecab による形態素解析では辞書に登録されていない語の原形は*として表示される。このため、解析した結果原形が*となった語については、最初に出現した形を原形として示している。

表 2 調査した全動画における出現頻度の集計結果

TF 値	原形	品詞
175,221	88	名詞
150,024	ww	名詞
85,406	一一	名詞
65,966	そらる	名詞
57,801	1	名詞
51,061	まふまふ	名詞
33,331	父祖	名詞
32,871	らま	名詞
27,029	4	名詞
20,349	する	動詞

次に、定義した DF 値を用いて描画したヒートマップを図 2 に示す。縦軸は DF 値の上位 10 語を順に並べた各語、横軸は動画が投稿されてからどの程度経ったコメントかを示す 4 つの日時区分 (1 日以内, 1 週間以内, 1ヶ月以内, それ以降) を表している。なお、縦軸に並ぶ語彙に括弧書きされている値が DF 値である。

表 3 カテゴリごとの出現頻度の集計結果 1
(a) ゲームカテゴリでの集計結果 (b) アニメカテゴリでの集計結果

TF 値	原形	品詞	TF 値	原形	品詞
65,426	ww	名詞	41,909	ww	名詞
15,142	一一	名詞	30,930	一一	名詞
4,241	88	名詞	30,915	1	名詞
4,105	する	動詞	23,398	88	名詞
2,480	うぼつ	名詞	18,800	4	名詞
2,168	かわいい	形容詞	4,079	する	動詞
1,841	さん	名詞	3,603	3	名詞
1,710	声	名詞	3,182	2	名詞
1,530	キヨ	名詞	2,398	かわいい	形容詞
1,378	1	名詞	2,205	可愛い	形容詞

表 4 カテゴリごとの出現頻度の集計結果 2

(a) 歌ってみたカテゴリでの集計結果 (b) 科学カテゴリでの集計結果

TF 値	原形	品詞	TF 値	原形	品詞
139,595	88	名詞	24,409	1	名詞
65,966	そらる	名詞	12,089	ww	名詞
51,060	まふまふ	名詞	11,085	2	名詞
33,331	父祖	名詞	9,333	3	名詞
32,864	らま	名詞	7,987	88	名詞
31,396	一一	名詞	7,938	一一	名詞
30,600	ww	名詞	7,612	5	名詞
15,492	飲む	動詞	7,538	4	名詞
9,445	うい	形容詞	5,051	9	名詞
8,003	俺	名詞	4,703	する	動詞

5 考察

5.1 出現頻度における考察

表 2 から表 4 の結果より、実験評価した 4 カテゴリ全てに共通して「w」を用いたコメントが出現頻度の上位に表れている。「w」は Murakami らが動画の評価指標として用いているように、動画やカテゴリに依存せず広く投稿されているコメントだと推測できる [9]。本実験でも先行研究で指摘されている事が裏付けられたと言える。

シンボル「88」についても 3 カテゴリに見られ、「w」と合わせてこれらのコメントは、カテゴリに依存せず投稿されているという結果であった。この「88」は、拍手を表すコメントであり、カテゴリの性質上歌ってみたカテゴリでより高い順位の頻度が示されているとともに、分析した他のカテゴリにおいても

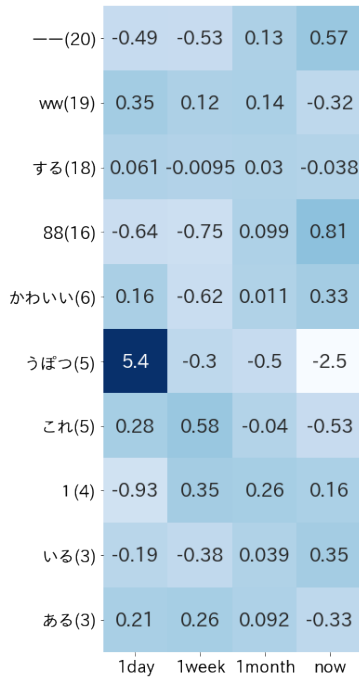


図2 DF 値の上位 10 語と投稿日時における時系列

高頻度で出現している。アニメカテゴリの集計結果では「88」以外にも数字のコメントが高頻度で投稿されているが、これは動画内で提示されている選択肢が 1 から 4 の 4 択問題へ答えるために投稿されていると考えられる。このため、「88」とは異なり動画の内容依存のコメントであると言える。選択肢のような数字本来の意味で投稿されたコメントでは、「44」や「444」のような同一の数字が異なる回数繰り返すコメントが多数回出現することは考えにくく、視聴者コミュニティの中で特別な意味・概念を付与されたコメントであるかの切り分けは、比較的容易に行えると考えられる。

他の語彙・シンボルに関して、neologd 辞書にも登録されていない語彙が多く、形態素解析が失敗している場合が有り、コメント特有の語彙を集めた用語辞書を用意するなどの対応が必要である。また、各動画について動画内の登場人物や動画投稿者と思われる人物の名前が多く投稿されている。こうした語も辞書に登録することで、コメントに用いられる語の抽出精度を向上できると考えられる。

5.2 コメント付与時間における考察

図2に示したコメント付与時間における時系列分析では、DF 値が高い語、つまり動画やカテゴリに非依存だと考えられるコメントは 1ヶ月以降に投稿された割合が非常に高いことがわかる。「w」や「88」も例に漏れず 4 列目の色が強く濃くなっている結果となっている。しかし、DF 値が 0 の語についても同様のヒートマップを作成すると同じ傾向を持つことが多く、DF 値のみで後から付与されるコメントほど動画やカテゴリに依存しない割合が高いとは判断できない。加えて、カテゴリごとについても図2と同様のヒートマップを描いた結果、アニメカテゴリや歌ってみたカテゴリで DF 値が 1 以上となっている語は特に 4 列目の色が濃くなっており、ゲームカテゴリと科学カテ

ゴリによく現れている語彙はアニメカテゴリや歌ってみたカテゴリほど 4 列目が濃くなっていないように見えた。

以上を踏まえ、図2の各行（語彙）をベクトルとみなしてクラスタリングを行い、各クラスタと DF 値・カテゴリの関係性の解釈を試みる。また、分析にあたり TF 値の上位 1000 語を対象とする。これは DF 値が 1 以上の語を全て含むだけでなく、DF 値が 0 となる語も含むため、DF 値を用いた比較がしやすくなるようにしている。ヒートマップを描くにあたり語彙ベクトルを正規化したが、この後にヒートマップの列にあたる各日時区分について標準化を行い、平均と分散を考慮したクラスタリングが行えるようにした。クラスタリングには k-means 法を用い、クラスタ数は調査カテゴリ数と同じ 4 とした。クラスタリングを実行し、クラスタごとの語彙をヒートマップに示したの中から、TF 値の上位 10 語を抽出した図を以下の図3と図4に示す。



図3 TF 値上位 1000 語のクラスタリング結果 1

図3および図4から、DF 値が高い語は各クラスタに分散しており、DF 値が高いと類似した時系列分布を持つとは一概には言えないことがわかる。

クラスタごとの時系列分布に注目すると、TF 値が全体的に低い語（クラスタ 1）・1day や 1week に投稿が偏っている語（クラスタ 2）・1month_jに投稿が偏っている語（クラスタ 3）・満遍なく投稿されている語（クラスタ 4）と集約されている。

クラスタごとに含まれる語に注目すると、クラスタ 3 では「俺」や「そらる」などの人物を指す名詞がいくらか含まれている。TF 上位 10 語以外の結果も見ると「可愛い」や「好き」などの形容詞がクラスタ 3 に含まれており、動画投稿者やユーザー自身についてのコメントの大半は動画が投稿されてから 1ヶ月以上後に付けられるものだと考えられる。これは歌ってみたカテゴリなどの動画投稿者にブランドがある動画について、動画投稿者が有名になっていくにしたがってコメントするユーザーが増えていった現象を示していると考えられる。クラスタ 4 については「ww」や「する」といった語彙が上位にある。これ

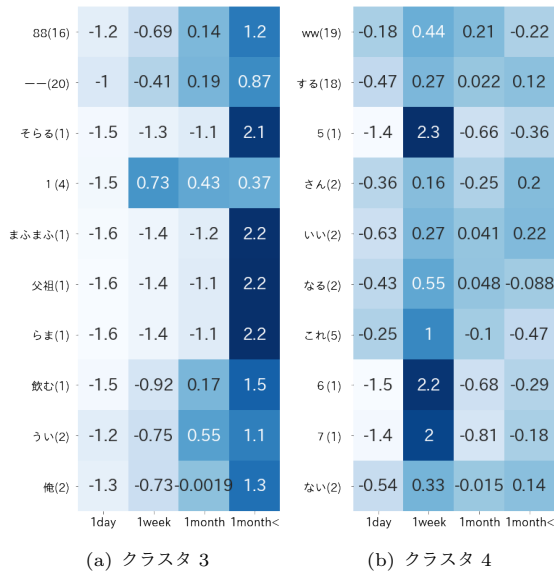


図 4 TF 値上位 1000 語のクラスタリング結果 2

らはカテゴリ非依存の語彙と考えられるため、時系列を問わず満遍なく付与されていると推測できる。なお、その場合日時区分の設け方から 1ヶ月以降（ヒートマップの 4 列目）の割合が高くなるようにも考えられる。しかし、一般に動画のコメント数は日時が経つにつれて減衰していくことを考慮すると、日時区分で分割した際に、ヒートマップの濃淡に差がない方が時系列を問わず満遍なくコメントが付与されていると考えられる。

次に、各クラスタとカテゴリの関連性を検証するため、クラスタごとの DF 値の出現の仕方をカウント、カテゴリ間の類似度を計算する。各カテゴリにおける DF 値の大小が、TF 値が上位となる語をどの程度含むかを表す。したがって、各カテゴリの DF 値を要素とするベクトルの類似度によって各クラスタの特徴とカテゴリの関連性を示せる。DF 値がカウントされている語について、その DF 値がどのカテゴリの動画によってカウントされたものを調べた結果が表 5 である。本研究において、DF 値は各動画について TF 値が上位 10 語である語に対して値を 1 加算した数量として用いているため、各カテゴリの DF 値の総和は 50 となる。

表 5 クラスタごとの DF 値の集計結果

カテゴリ	ゲーム	アニメ	歌ってみた	科学	クラスタの語彙数
クラスタ 1	2	6	1	4	73
クラスタ 2	4	0	5	7	106
クラスタ 3	18	32	30	13	357
クラスタ 4	26	12	14	26	464

表 5 の結果を元に、カテゴリ間の類似度を計算する。表 5 の 1 列目から 4 列目をそれぞれベクトルとみなし、これらのベクトルの類似度をコサイン類似度を用いて計算する。この際、クラスタごとに含まれる語彙数が違うため、各行において DF 値をクラスタの語彙数で割った値について類似度を求めることとする。

各カテゴリにおける DF 値のベクトルの類似度を表 6 に示す。

表 6 各クラスタに付けられた DF 値の類似度

	ゲーム	アニメ	歌ってみた	科学
ゲーム	1	0.75	0.89	0.93
アニメ	—	1	0.74	0.68
歌ってみた	—	—	1	0.78
科学	—	—	—	1

表 5 から、ゲームカテゴリと科学カテゴリは確かに似た DF 値の付けられ方がされており、表 6 によってその類似性が示されている。一方、アニメカテゴリと歌ってみたカテゴリについてはコサイン類似度が 0.74 となり、ゲームカテゴリと科学カテゴリほどの類似性はみられないことがわかる。

5.3 動画のタイムラインにおける考察

動画のタイムラインにおける時系列分析においては、既にコメントが多く付与されている位置に上乘せされるようにコメントが付与されていくのであれば、既に付与されているコメントが新たなコメントに影響を与えており、バンドワゴン効果が観測されたといえる。

これを検証するため、動画が投稿されて十分経った後に投稿されるコメントは、動画の盛り上がっているシーン（すでに多くのコメントがついているシーン）に上乘せするようにつけられることが多いと考えた。それぞれの動画について、各コメントの投稿日時とタイムライン上での位置を棒グラフに示すことで、定性的な結果を示す。投稿日時についてはここまでの日時区分と同様の 4 区分を用い、動画のタイムラインについては全体を 10 分割して示す。図 5 では、ある動画について、横軸をコメントのタイムライン上での位置、縦軸をコメントの頻度としたヒストグラムを示している。(a) は動画内の全コメントについて投稿位置を示した棒グラフであり、(b) は図 2 にて高 DF 値を持つとわかった「ー」「w」「8」を含むコメントについて投稿位置を示した棒グラフである。しかし、動画の全コメントを対象にタイムライン上での投稿位置を分析しようとしたところ、動画のタイムラインとコメント投稿時間に目立った関係は見られなかった。

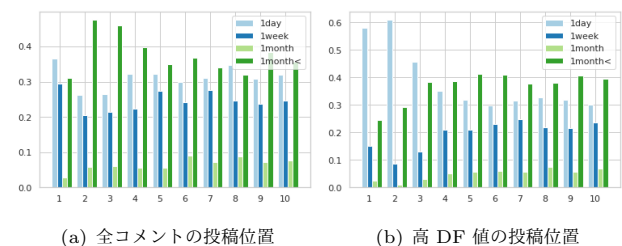


図 5 ある動画におけるコメントの投稿位置

6 まとめ

本研究では、ビデオのタイムライン上に重畳表示されるコメントの分布を明らかにする分析を行う第一段階として、コメントに使用されている語彙の抽出を実装し、幾つかのカテゴリの

ビデオのコメントを解析した。形態素解析をベースに、コメント付与のヒューリスティクスを正規表現のルールとして埋め込むことで、時系列分析に使用できる程度の精度で語彙を抽出することができたと言える。

コメント付与時間における時系列分析の結果、投稿日時の時系列分布はDF値に依るとは言えず、各語彙のクラスタリング結果から、カテゴリという動画の性質を示す分類に応じてコメントの時系列的振る舞いが異なることが示唆された。また、動画のタイムライン上における投稿位置については、コメントのDF値に応じて分布が変わることが示唆された。これはカテゴリに応じたフィルタリング機能の提案に繋がる点で意義深いといえる。

今後は、カテゴリごとの動画の特徴を加味した分析、具体的にはカテゴリごとに見られる再生時間の違いなど、動画のタイムラインにより注目した分析を行うことで2時系列間の関係性を明らかにしたい。加えて、株式会社ダウンゴが告知しているように、現在ではカテゴリの仕様が変更されている[10]。具体的には2019年以前は大カテゴリと小カテゴリからなる構成であったが、ジャンルと利用頻度の高いタグからなる構成へと変更されている。今後はこの仕様変更に適応したデータを用いた分析にも取り組みたい。

謝 辞

本研究はJSPS科研費JP16H02904の助成を受けたものです。本研究では、国立情報学研究所のIDRデータセット提供サービスにより株式会社ダウンゴから提供を受けた「ニコニコ動画コメント等データ」を利用した。

文 献

- [1] 株式会社ダウンゴ. "ニコニコ動画(く)". ニコニコ動画. https://www.nicovideo.jp/video_top?header, (参照 2020-01-07)
- [2] 前島紘希, 中村聡史, 土屋駿貴, 大野直紀. (2016, March). 印象に関する検索意図を考慮したサムネイル動画自動生成手法の提案. 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016) (Vol. 4, pp. 1-7). 一般社団法人電気情報通信学会.
- [3] 平澤真大, 小川祐樹, 諏訪博彦, 太田敏澄. (2013). ニコニコ動画のログデータに基づくソーシャルノベルティのある動画の発見手法の提案. 情報処理学会論文誌, 54(1), 214-222.
- [4] 齊藤義仰, 村山優子. (2011). 視聴者コメントを用いた広告動画挿入タイミング決定アルゴリズムの提案と評価. 情報処理学会論文誌, 52(2), 520-528.
- [5] 佃洗撰, 中村聡史, 田中克己. (2011). 視聴者のコメントに基づく動画検索および推薦システムの提案. 第19回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2011) 論文集, 78-83.
- [6] 山本岳洋, 中村聡史. (2013). 視聴者の時刻同期コメントを用いた楽曲動画の印象分類. 情報処理学会論文誌データベース (TOD), 6(3), 61-72.
- [7] 磯貝佳輝, 齊藤義仰, 村山優子. (2011). 視聴者コメントを用いた動画検索支援のためのダイジェスト動画作成アルゴリズムの検討. 情報処理学会第73回全国大会 (IPSJ2011), 2011(1), 347-348.
- [8] 株式会社ダウンゴ (2019): ニコニコ動画コメント等データ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.3.1>
- [9] Naomichi Murakami and Eisuke Ito. 2011. Emotional video

- ranking based on user comments. In Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (iiWAS '11). Association for Computing Machinery, New York, NY, USA, 499-502. DOI:<https://doi.org/10.1145/2095536.2095639>
- [10] 株式会社ダウンゴ. "【告知】動画のランキングおよびカテゴリの仕様変更". ニコニコインフォ. <https://blog.nicovideo.jp/niconews/103407.html>, (参照 2020-01-07).