

F-002

共感度検索のための動画アノテーション分析

村上直至[†] 伊東栄典[‡]九州大学 [†]大学院システム情報科学研究府 [‡]情報基盤研究開発センター
〒812-8581 福岡市東区箱崎 6-10-1E-mail: [†]2IE10081E@s.kyushu-u.ac.jp, [‡]ito.eisuke.523@m.kyushu-u.ac.jp

あらまし 近年, YouTube やニコニコ動画などの利用者投稿型の動画サービスが人気である。動画には作成者によるタイトル名の他に, 視聴者による各動画へタグ付けや動画コメントが付与される。ニコニコ動画では視聴回数やマイリスト登録数等の付属情報も取得できる。これら多数の視聴者からの付加アノテーション情報は, 視聴者の動画に対する共感を含むため, folksonomy として動画を特徴付けるマイニング資源に利用可能である。共感度を抽出・定量化できれば, 動画のランキング・推薦・分類に有効である。本研究では, ニコニコ動画を対象として利用者投稿型の動画コンテンツに付随するタグやコメントの分析を行った。

キーワード Web 情報サービス, ニコニコ動画, アノテーション, フォークソノミー

Movie annotation analysis for sympathy based ranking

Naomichi Murakami[†] and Eisuke Ito[‡]† Grad. School of ISEE, ‡ Research Institute for IT, Kyushu University
Hakozaki 6-10-1, Higashi-ku, Fukuoka, 812-8581 JapanE-mail: [†]2IE10081E@s.kyushu-u.ac.jp, [‡]ito.eisuke.523@m.kyushu-u.ac.jp

Abstract A lot of movie services are provided on the web such as YouTube, and Nicovideo. Nicovideo is a popular movie uploading and streaming service in Japan. Movies are given not only metadata such as title, creator, upload date and movie length by the creator, but also a lot of annotation such as tags and comments by viewers. Nicovideo system provides number of replay and my-lists mark count, additionally. Those annotations may be used as folksonomy based data mining, because viewer may present his/her sympathy in annotation. In this paper, we analyzed annotations. We hope that sympathy analysis from movie annotation may can use for another ranking method.

Keyword Web information service, Nicovideo, annotation, comment, folksonomy

1. はじめに

近年, YouTube やニコニコ動画[1][2]などの利用者投稿型の動画サービスが人気である。動画には作成者によるタイトル名の他に, 視聴者による各動画へタグ付けや動画コメントが付与される。ニコニコ動画では視聴回数やマイリスト登録数等の付属情報も取得できる。

ニコニコ動画は, 独特なコメントシステムによって国内で人気を博している, 視聴者投稿型の動画コンテンツサービスである。動画へのコメント投稿は, コンテンツの再生時間軸に沿って情報が保存され, 視聴者同士の擬似的な時間共有を可能にした。ニコニコ動画のコメントシステムでは, 多数の人間が, 多数のアノテーションを, それも時にはコンテンツの1秒に対して十数件もの密度で付与している。過去, 動画に対しこれほど多量

なアノテーションが付与されたことは無かった。

2010年3月末現在, ニコニコ動画はコメント情報を積極的に検索システム等に利用していない。代わりに, 利用者により動画に付けられたタグ情報を用いてコンテンツを分類し, 「再生回数」や「マイリスト登録数(お気に入り数)」といったコンテンツの人気度によるランキングサービスで各動画の重みを決め, 動画検索を支援している。

この, 大人数が視聴する動画は視聴価値が高いという手法では, 小~中人数が深く楽しむコンテンツに対し良い評価ができない。万人受けしないコンテンツは, 興味を持つ視聴者数が少ないため視聴回数が上位にならず, 上位でないために視聴者がやはり増えないという悪循環に陥る。ニコニコ動画では1日に数千件の動画が投稿されるため, 万人受けしない動画はすぐに数多くの動画コンテンツの海に沈んでしまう。これを避けるには,

別の動画評価尺度が必要である。

本研究では、多数の視聴者からの付加アノテーション情報は視聴者の動画に対する共感を含むため、folksonomyとして動画の特徴付けるマイニング資源に利用可能と考えている。アノテーションから視聴者が何に共感しているのかを汲み取り、視聴者の共感度を抽出・定量化できれば、動画のランキング・推薦・分類に有効であろう。

本稿では、ニコニコ動画を対象として利用者投稿型の動画コンテンツに付随するタグやコメントの分析結果を報告する。

2. ニコニコ動画

ニコニコ動画は2006年12月12日に運営を開始した、視聴者投稿型の動画配信サービスである。運営開始から3年で350万件を超える動画が投稿されている。会員数は一般会員が1410万人、有料会員数は70万人に上る[2]。

類似サービスのYouTubeと比べたニコニコ動画の最大の特徴は独自のコメントシステムである。動画のある再生時点に投稿されたコメントは、コメント投稿者の再生時間に、動画再生画面上を右から左に横切るように表示される。この仕組みは、異なる場所と時間で動画再生していながらも、まるで他の視聴者と一緒に動画を観ているかのような感覚を得る演出効果がある。

2.1. 動画に付与されるアノテーション

ニコニコ動画へ投稿された動画には、投稿者アカウント、動画ID、タイトル、説明文、投稿日時、タグ、サムネイル画像、動画の長さ、コメント、再生回数、マイリスト登録数のデータが付与される。このうちタイトル、説明文、投稿日時、タグの一部、コメントの一部、サムネイル画像、動画の長さ情報は投稿者が決定する。図1にアノテーションの概要を示す。

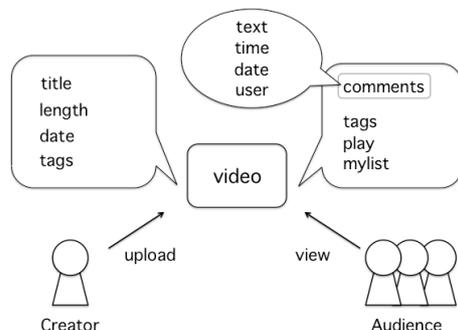


図1 ニコニコ動画におけるアノテーション

2.2. 動画の検索

サイト内の動画にはキーワード、タグ、キーワードを含むタグの3つの観点から検索を行うことができる。キーワード検索は、動画投稿者が付与するタイトルと説明文から検索絞込みができ、見たいものが最初から分かっている場合に有用である。タグ検索では、動画投稿者や視聴者が動画に付与したタグが検索対象となる。視聴者は動画への自由なタグ付けができるため、タグには集合知による情報が付与されているため、キーワード検索よりも柔軟に絞り込める利点がある。

ニコニコ動画では、検索の適合集合を、いくつかの手法に従ったランキングとして検索者に返す。ランキング手法に関する詳細は次節で述べる。

3. 「音楽」カテゴリ動画

ここでは、分析の対象とした「音楽カテゴリ」の動画について述べる。

全動画(350万件以上)の分析は困難であるため、本研究では「音楽」タグが付与された動画を対象とした。ニコニコ動画では、代表的なタグが、動画の大分類となるカテゴリタグに選ばれており、「音楽」は最初のカテゴリタグである。動画投稿者は、自分の動画を分類されやすくするため、カテゴリタグをロック状態で(他者が編集出来ない状態で)付与することが多い。

2009年10月28日より11月5日までの間、「音楽」タグが付与された動画373,265件について、アノテーションを取得した。最初に「音楽」タグを持つ動画の動画IDを取得し、各動画に付随するアノテーションを収集した。アノテーション収集にはニコニコ動画から公式に提供されているAPI[3]を利用した。動画IDリストの収集とアノテーション収集に時間差があるため、アノテーション取得の際に「音楽」タグが消えてしまっていた動画もあった。

373,265個の動画のうち5,828個には「音楽」タグがアノテーション取得時には消えていた。317,751件の動画にロックされた「音楽」タグ、49,686件の動画にロックされていない「音楽」タグが付けられていた。

4. タグの分析

ニコニコ動画では、一つの動画あたり最大10個のタグが付与できる。タグには、投稿者がロックする(他者による編集不可とする)タグと、一

般視聴者が自由に付与する非ロックタグがある。

4.1. 全タグの分析

収集した動画集合には、一意なタグが 356,362 個存在した。このうち 71%にあたる 251,540 個のタグは「音楽」タグと一度しか共起しなかった。一度だけ共起したタグは、他の視聴者をうならせるネタ的意味合いの強いタグや、付与された動画固有の説明文となっているタグが散見された。これらのタグは検索における効果は薄いものの、動画コンテンツを盛り上げる要素となっている。

図1はタグの順位と出現頻度図(各軸は対数軸)である。Zipfの法則にほぼ従う形となっている。

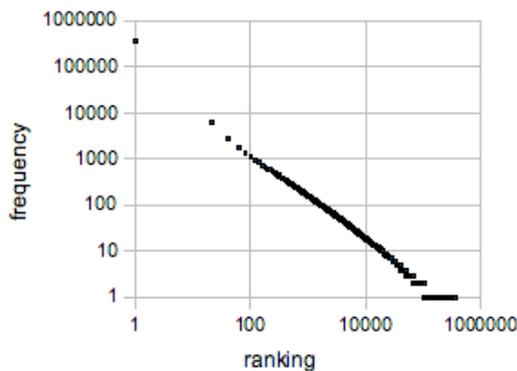


図1 タグの順位頻度図

表1に上位10位のタグと出現頻度を示す。1位の「ニコニコムービーメーカー」と第5位の「投稿者コメント」は、ニコニコ動画のシステムが自動付与するタグである。

表1 出現頻度上位10位のタグ

順位	出現頻度	タグ
1	49007	ニコニコムービーメーカー
2	34236	VOCALOID
3	31002	作業用 BGM
4	30255	初音ミク
5	29689	投稿者コメント
6	27950	洋楽
7	17100	東方
8	12993	ミクオリジナル曲
9	11943	もっと評価されるべき
10	9826	邦楽

ニコニコムービーメーカーは、2008年公開の静止画と音声を組み合わせて動画ファイルを作る無償ツールである。手軽に動画を制作できるため人気がある。このツール経由の投稿動画には「ニコニコムービーメーカー」タグが、ロック状態で自動付与される。

「ニコニコムービーメーカー」が高頻度で出現した理由は、音楽タグを持つ動画を解析対象にしたためであろう。音楽動画は音声の主であり画面は重要でない。同ツールが音楽を主題とした動画を制作するのに向いていたために、高い頻度で出現したと考えられる。

5位の「投稿者コメント」タグは、動画投稿者がロックしたコメントを入れた場合に自動付与される。なお、自動付与タグは投稿者が自発的にタグロックを解除することで消去できる。

自動付与タグでない高出現頻度タグには「VOCALOID」「初音ミク」「ミクオリジナル曲」等のボーカロイド関連タグがある。7位の「東方」はゲームソフトのシリーズ名で、特にBGMの評判が高い。これらはニコニコ動画のサービス開始から2009年後半までの、音楽タグのついた動画の特徴を表している。

4.2. ロックタグの分析

次に動画投稿者がロックしたタグに注目した。一意なロックタグ数は66,582個であった。一つの動画のみ付与されているロックタグは40,527個、60.9%であった。図2にロックタグのFR図を描いた。この図もZipfの法則にほぼ従っている。

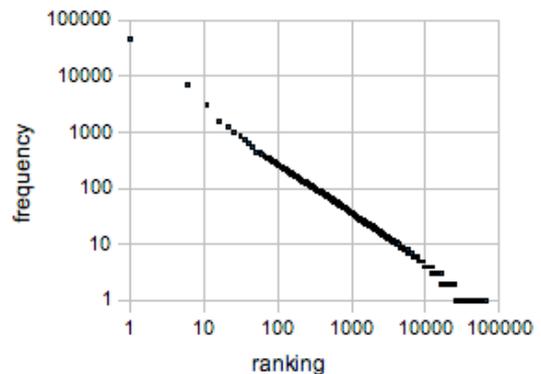


図2 ロックタグの順位頻度図

表2 出現頻度上位10位のロックタグ

順位	出現頻度	タグ
1	44221	ニコニコムービーメーカー
2	27059	投稿者コメント
3	12872	初音ミク
4	9422	作業用 BGM
5	7123	VOCALOID
6	6989	東方
7	6895	洋楽
8	6762	@BGM 推奨
9	3766	ミクオリジナル曲
10	3086	KAITO

投稿者がロックしたタグには、検索に有用なタグに加え、先に述べたネタ的意味合いの強いものである、その動画固有の約束事や冗談などがあつた。

4.3. 非ロックタグの分析

最後に、ロックをかけられていないタグを分析した。図3に非ロックタグのFR図を示す。この図もZipfの法則に従っている。

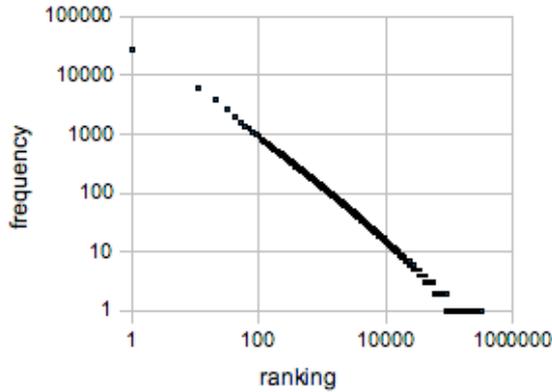


図3 非ロックタグの順位頻度図

表3 出現頻度上位10位の非ロックタグ

順位	出現頻度	タグ
1	27113	VOCALOID
2	21600	作業用 BGM
3	21053	洋楽
4	17383	初音ミク
5	11771	もっと評価されるべき
6	10111	東方
7	9445	邦楽
8	9227	ミクオリジナル曲
9	7140	PV
10	6652	VOCALOID カバー曲

9位と第10位を除く他のタグは、表1の上位10位に入っている。ここで特徴的なことは、「もっと評価されるべき」と「邦楽」は、ほとんど非ロック状態で出現していたことである。「洋楽」タグのロック率が24.7%であるのに対して、「邦楽」タグのロック率が4.9%であったことは興味深い。

非ロックタグは視聴者によって付与される。全視聴者により編集可能であるため、削除されることもある。その中で、削除されずに残ったタグは、その動画への視聴者の思い、共感が含まれていると考えられる。例えば、「もっと評価されるべき」タグは、良質動画にも関わらず再生数等の評価が低い、あるいは存在自体認知されていないと感じられた場合に、視聴者側から自然発生的に付与さ

れていると考えられる。

4.4. 再生数, コメント数, マイリスト登録数, 動画長の分析

収集した全動画からランダムに1,000件を抽出し、それらの動画再生数, コメント数, 動画長について順位-頻度をプロットした。動画長は秒を用いている。

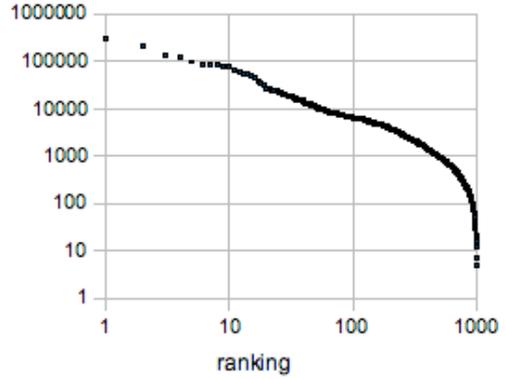


図4 順位-再生数

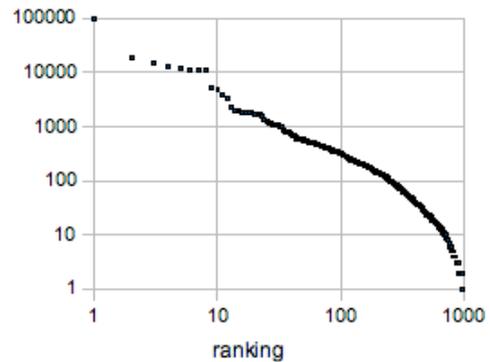


図5 順位-投稿コメント数

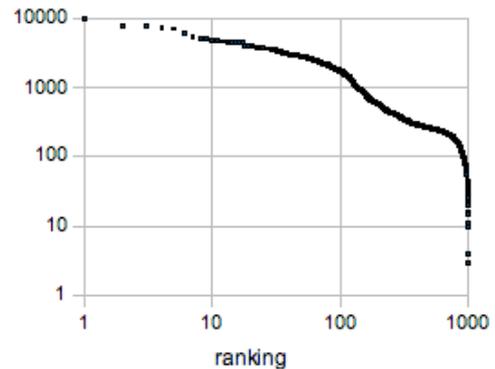


図6 順位-動画長

図4,5はZipfの法則の拡張であるLavaletteの法

則[4]に近い。順位-頻度のプロットが Lavalette の法則に近くなる例として、はてなブックマークにおける、URL の被ブックマーク数とその順位がある[5]。はてなブックマークは、話題となっている URL を利用者に推薦するサービスを提供している。文献[6]では、多数の視聴者がチェックした動画を自分もチェックするという視聴者の行動パターンが広く存在するとき、平均値付近の頻度が増加し、FR 図は Lavalette の法則に近い軌跡を描く可能性があるとして指摘している。

動画長は、81.5%の動画が10分以下で、4分以上5分未満の動画が最も多い。一般的な楽曲の長さは5分程度であり、ニコニコ動画でも音楽タグが付与された動画の多くは一曲分の長さを持つことが分かる。

最後に、再生数、コメント数、マイリスト登録数、動画長(play, comment, mylist, Length)について相関係数を算出した。結果を表4に示す。

表4 相関係数

play	play			
comment	0.73	comment		
mylist	0.88	0.71	mylist	
length	0.01	0.05	0.02	length

動画長は他の値と関係が無いことが分かる。つまり、長時間動画でも再生が多い場合もあり、また動画が長くてもコメントは増えるとは限らない。一方、再生回数、コメント数、マイリスト登録数は正の相関を持つ。特に再生数とマイリスト登録数との相関は9割に近い。音楽カテゴリ全体での再生回数、マイリスト数の降順ソートを比べたところ、上位30件の動画の内、23件が同一の動画であった。

コメント数と再生数の相関は7割を超えたが、マイリスト数と再生回数が見せた相関係数よりは若干低い。コメントをするべきでは無い、静かな気持ちで閲覧したい動画や、たくさんコメントを投稿して画面を賑やかに演出することが目的の動画などが存在し、相関を崩したからだと考えられる。

5. 視聴者コメントの分析

ここでは、ニコニコ動画最大の特徴であるコメントの分析を述べる。

5.1. コメント内の共感表現

動画に対して視聴者が付与するコメントには、

しばしば視聴者が動画に向ける感情の切れ端が含まれる。低い再生回数の動画であっても、視聴者が楽しそうに動画を閲覧している様子はコメントから伝わってくる。

コメントに出現する単語には、動画に特定の単語が頻出することもあるし、分野(カテゴリ)に固有の単語もある。また、ニコニコ動画内で一般的な文字列もある。たとえば、文字 w を用いたコメント高い頻度で出現する。

ニコニコ動画には「warai」の略である文字「w」を使って動画へのプラスの感情を表現する文化がある。コメント内の w には、もとの意味である「笑い」だけでなく、気分が高揚している状態全般を表す語として多用されている。

そのため、視聴者コメントから視聴者の共感度を図ることが可能であると考えている。さらに、共感度を使うことで動画推薦や動画ランキングを実現できると考えられる。今回は、共感度分析の前段階として、コメント内の「賞賛」や「笑い」などの感情を読み取ることにした。

5.2. コメント内の w 度

動画の共感度解析の一例として、文字 w を用いた解析を行った。先に述べたように、笑いなどの感情が文字 w を用いたコメントで表されている。そこで、視聴価値の高い(おもしろい)動画には、共感を表す w が頻出するという仮説を立て、現在その検証を行なっている。

そこで、文字 w を用いた動画価値の新しい評価手法を提案し、提案手法を用いたランキングシステムを試作した。このシステムでは、多分に興味深い動画ランキングを実現できた。その上位に並ぶ動画群は決して万人受けしない、しかし視聴者のほとんどがその動画を楽しんでいる様子が、確認できるものであった。

6. 関連研究

ニコニコ動画は、常に新しいサービスが実験的に生み出されては運用が始まる変化の激しい場であるために、自発的に研究・解析を行っていた利用者も多い。2009年1月には、自発的な研究を行っていた研究者が、ニコニコ動画データ分析研究発表会[7]を開催した。末永匡は榊剛史らの論文[8]に倣い、値を用いた単語間の関連度の指標を使って、ニコニコ動画の folksonomy なタグ同士の関連度を求めた[9]。文献[10]はタグ毎の動画総再生数は対数正規分布に従うとして、中央値から総

文 献

再生数を推定する手法を提案したものである。タグ同士の類似度とタグ毎の動画再生数の分布から、任意の動画の再生回数が推定できれば、推定値と実測値の類似度を用いてもっと評価されるべき動画を自動検出できる可能性がある。

動画に付随するアノテーションは、動画の内容検索に重要な意味を持つ。Harald SackらはMPEG-7のアノテーション構造を用いたインタフェースの例を示した[11]。

また、にこぐらふ [12]は、我々の手法[6]と同様に文字 w に着目した、動画のどの部分が面白いのかを発見するサービスである。リクエストがあった動画のみ最新 500 件のコメントを取得、時間を横軸に文字 w の数と縦軸にグラフを描画して、動画毎の面白いポイントを視覚的に把握できるようにしている。2008年3月のサービス開始からまもなく2年になる同サイトには現在、345件の動画が登録されている。独自の公開されていない計算式による「にこ率」も定義しており、にこ率による登録動画ランキングも提供している。

中村貴洋らは、ニコニコ動画のコメント情報をコンテンツへのアノテーションと捉え、動画毎の見所を自動検出する実験を行った[13]。コメントの集中する箇所のサムネイル画像を並べ、視聴者の印象に残ったと予想されるシーンをマンガ的に表示するシステムを構築している。

7. おわりに

本論文ではニコニコ動画を対象に、視聴者から付与されたアノテーションの分析をおこなった。具体的には、動画のタグと、動画に寄せられたコメントを分析した。

今後の課題は3つある。一つ目はタグ利用である。タグ間の階層関係による動画の階層関係を考えたい。また、タグとコメント内の単語関係を調査したい。

二つ目はコメントの詳細の分析である。現在の所、コメント内の w しか解析できていない。他の文字・単語について、出現頻度や共起関係を調査したい。

最後は視聴者の信頼度を評価したい。単語の頻度解析だけでは、コメントを投稿した視聴者の価値を考慮していない。気持ちをうまく代弁する良質な視聴者のコメントは、他のコメントに比べ重要視されるべきである。これらの問題を解決していきたい。

- [1] ニコニコ動画, <http://www.nicovideo.jp/>,(accessed 2010.2.1).
- [2] ニコニコ動画:From Wikipedia, the free encyclopedia., <http://ja.wikipedia.org/wiki/ニコニコ動画>, (accessed 2010.2.1).
- [3] getthumbinfo, <http://ext.nicovideo.jp/api/getthumbinfo/>.
- [4] I. Popescu: On a Zipf's Law Extension to Impact Factors, *Glottometrics* 6, pp.83-93, (2003)
- [5] 百田信, 伊東栄典: ソーシャルブックマークに基づく情報発見, 電子情報通信学会 第19回データ工学ワークショップ (DEWS 2008), pp.11-15 (2008.3)
- [6] 村上直至, 伊東栄典: 視聴者コメントに基づく動画検索手法, 電子情報通信学会 2010年総合大会, D-4-2 (page D-24), Mar. 2010.
- [7] CodeZine: 膨大なデータを分析して見えてくること, ニコニコ動画データ分析研究発表会, <http://codezine.jp/article/detail/3516>, (accessed 2010.1.18)
- [8] 榊剛史, 松尾豊, 石塚満: Web上の情報を用いた関連語のシソーラス構築について, 自然言語処理, Vol.14, No.2, pp.3-31 (2007.4)
- [9] 末永匡: χ^2 乗値を関連度としたニコニコ動画関連タグネットワークの解析, <http://www.hotdocs.jp/file/178654>, (accessed 2010.1.18)
- [10] myrmecoleon: ニコニコ動画におけるタグ別動画群の総再生数推定の方法について, <http://www.hotdocs.jp/file/178659>, (accessed 2010.1.18)
- [11] Harald Sack, Jorg Waitelonis: Integrating Social Tagging and Document Annotation for Content-Based in Multimedia Data, SAAW2006 - 1st Semantic Authoring and Annotation Workshop Athens, GA, USA, (2006.11)
- [12] にこぐらふ, <http://www.nicograph.jp/>.
- [13] 中村貴洋, 青木秀憲, 宮下芳明: マンガ的手法を用いたニコニコ動画ナビゲーション, ヒューマンインタフェース学会研究報告集, vol.10, pp.103-110 (2008.)