

Bilibili 動画サービスにおける感情コメント分析

A study of emotional user comments analysis for bilibili.tv

呉 沢臣¹ 伊東 栄典²

Zechen Wu¹, and Eisuke Ito²

¹九州大学大学院システム情報科学府

¹Dept. of ISEE, Kyushu University

²九州大学情報基盤研究開発センター

² Research Institute for Information Science, Kyushu University

Abstract: On user generated contents sites, user's comments include the user's impression to the content. The authors believe that user's comments can be used as the data mining resource to evaluate the contents. In this paper, the authors focus on bilibili.tv, a Chinese video sharing site, and analyze user's emotional comments on the site. They also show co-relation between comments and popularity such as number of replay and bookmark.

1 はじめに¹

近年, YouTube (youtube.com), ニコニコ動画 (nicovideo.jp) などの利用者投稿型動画共有サービスが人気である. 中国でも優酷 (优酷 youku.com) や土豆 (tudou.com), 搜狐 (sohu.com) および哔哩哔哩動画 (哔哩哔哩动画 bilibili.tv) が人気である. これらのサイトには膨大な動画が蓄積されている. 膨大な動画から, 利用者が求める面白い動画を探すためには検索システムが必要である.

我々は日本のニコニコ動画を対象に, 動画への視聴者コメントに基づく動画ランキング手法を提案してきた[1,2]. ニコニコ動画は視聴者投稿型の動画共有サービスで, 独自のコメントシステムで人気を博している. 本研究では, 中国の動画共有サイトである哔哩哔哩動画 (以下 bilibili と表記) 対象に, 視聴者コメントに基づく動画の品質評価を検討している.

Bilibili は, AcFun の会員である@bishi 氏が 2009 年 6 月 26 日に設立したファンサイト「Mikufans」を前身としている[3,4]. Mikufans は 2010 年 1 月 24 日から現在のサイト名「哔哩哔哩 (bilibili)」に改名され, 現在まで続いている. Bilibili は動画の実データを自サイトに保存しておらず, 別の動画サイトが保持する動画を使い, それに利用者向けのメタデータ (主に中国語) 提供している. また他サイトにある

動画再生に同期して, 弹幕と称する視聴者コメントを動画画面上に表示する機能や, 動画についての利用者のお気に入り情報の保持, およびファン向けのグループ作成機能などを提供している. これらの機能から, Wikipedia[3,4]には bilibili は寄生型サイトと表記されている.

近年, 日本の近代文化であるゲーム・漫画・アニメ, J-POP・アイドルなどのポップカルチャーをクールジャパン推進の一部として海外に積極的に展開する動きがある. 既にゲーム・漫画・アニメ・J-POP・アイドル等は海外に根強いファンが居るとの報告もある[5]. Bilibili も日本発の文化を受けたサイトの 1 つであると言える. 実際, Bilibili は日本のニコニコ動画およびニコニコ動画から発した Vocaloid 文化に強く影響されている. 前身サイトの Mikufans の Miku は初音ミクであるし, またサイトの構成はニコニコ動画に類似している. 動画再生に同期しての動画画面上への弹幕コメント表示は, ニコニコ動画のコメント表示と同じ機能である. 更に, ニコニコ動画サイトへの投稿動画を bilibili 内で表示する場合も多い. なお, Bilibili にはニコニコ動画以外にも, 米国や韓国系の動画も多い.

我々は, 動画に視聴者が付与するコメントのうち, 感情的なコメントを定量化し, その値を動画検索時のランキングに用いる事を検討している. 本論文では Bilibili を対象に, 視聴者コメントの統計的な分析を行った. Bilibili における感情コメント「233」について述べ, 再生数やブックマーク数などの人気度と感情コメント「233」の相関を調査する.

² 連絡先: 九州大学情報基盤研究開発センター,
〒812-8581 福岡市東区箱崎 6-1 0-1
Email: ito.eisuke.532@ml.kyushu-u.ac.jp

2 Bilibili 動画タグとコメント

2.1 Bilibili のメタデータ

最初に, Bilibili に投稿されている動画メタデータを 2013 年 11 月から 2014 年 1 月の間に収集した. Bilibili では動画に「AV ID」と呼ぶ識別子を付与する. AV ID は文字 av と数字で構成される. そのため, 数字部分を 1 から 99 万まで 1 つずつ増やしてアクセスした所, 440,849 個の動画についてメタデータ (HTML 形式) を収集できた.

図 1 に示すように 1 つの動画のメタデータには, 題名, 投稿者 (登録者名) の他に, 動画再生回数 (播放), コイン数 (硬币数量), マイリストおよびブックマーク数 (收藏), 弹幕コメント数 (弹幕) の情報が有る. 動画再生に合わせて画面上を流れるコメントを「弹幕」と名付けられている.

項目	値
播放 (播放)	42336
收藏 (收藏)	2307
弹幕 (弹幕)	1674
积分 (积分)	1715
硬币 (硬币)	551
当前观众 (当前观众)	1

図 1 Bilibili での動画利用状況情報 (av689970)

図 2 に動画のタグの例を示す. タグは視聴者により動画へ付与されている. ニコニコ動画と同様に, bilibili でも最高 10 個のタグを動画に付与できる. 図 2 の動画は, ニコニコ動画で「sm21443197」の識別子を持つ動画であるため, その情報がタグに記載されている.

タグ
sm21443197
音楽: Mitchie M, MVディレクション, MMD: cort, コンポジティング: TOHRU MITSUHASHI, 补起来

図 2 Bilibili でのタグ情報 (av689970)

2.1 タグの個数と頻度

集めた全動画メタデータからタグを抽出した. 一意なタグは 345,140 個あった. 1 つの動画にしか付与されていないため, 出現頻度が 1 回のタグは 25 万個あった. これは全体の 72.4%である. 表 1 に, 頻度 1 ~ 5 回のタグ個数を示す.

表 1 一意なタグの個数

頻度	タグ数	割合
1 回	249,825	72.4%
2 回	34,013	9.9%
3 回	14,595	4.2%
4 回	8,579	2.5%
5 回以上	38,128	11.0%
総数	345,140	100.0%

2.3 Bilibili の視聴者コメント

Bilibili には 2 種類の視聴者コメントが有る. 一つは動画再生に合わせて画面上に流れるコメントである. これは「弹幕」と呼ばれている. 弹幕コメントは匿名で, 投稿者を特定できない. もう一つは会員名付きのコメントである. これを本稿では「長コメント」と呼ぶ. 図 3 に示すように長コメントの例を示す.

コメント
抱枕被褥 (7/21)
2014-02-13 17:10
2014-01-19 19:47
2014-01-08 01:49
2014-01-07 11:22
2013-12-14 22:10
2013-12-13 11:51

図 3 Bilibili での長コメント (av689970)

3 感情コメント「233」

我々はコメントに含まれた感情を定量化することで, 視聴者の感情に基づく動画ランキングを作りたいと考えている. その第一歩として, 弹幕コメントおよび長コメントにおける, 感情コメントを分析した.

3.1 コメント数

集めた動画メタデータからコメント情報を抽出した. 表 2 にコメントの件数を示す. 文字数が多い長いコメントも, 1 文字で終わる短いコメントも 1 件

と数えている。

表 2 視聴者コメント数

種類	コメント数	平均
弹幕	72,622,153	167.4
長コメント	5,173,862	11.92
総数	77,796,015	179.33

3.2. 弹幕における感情と「233」

視聴者が付与するコメントには、視聴者が動画に対して保持する感情が含まれている。本研究が対象とするコメント、特に弹幕コメントは、短いフレーズや絵文字が多い。

実際、Bilibili の動画を調査したところ、面白い動画には「233」や「2333333」のように 3 が多く続く弹幕コメントが多く与えられている。233 は中国語圏でのネットスラングで、大爆笑の意味で用いられている。この由来は、中国の BBS の「猫扑」における笑顔の絵文字である。猫扑では 233 番の絵文字が笑顔であった。そこから 233 が笑いの感情を表す文字として多用されるようになった。23333 のように 3 が多くなると笑いの感情が強い。

233 の用法は、ほぼ日本語のネットスラング「www」と同じである。日本の「www」は笑い(warai)の頭文字が由来である。日本のニコニコ動画では、w が多いほど笑いの感情が強い事を示している。

感情コメント分析の最初として、笑いの感情を表す「233」を分析する。ここで、弹幕コメントに「233」が出現する動画を 233 動画と呼ぶことにする。全 440,849 個動画の中で、233 動画の数は 134,411 個である。30.5%の動画が 233 動画である。

表 3 233 動画の統計

種類	動画数
全動画	440,849
233 動画	134,411

文字列「233」を含む弹幕コメントは 1,403,180 個である。233 動画では、平均 10.44 回「233」が出現する。

表 4 233 コメントの統計

	233 コメント数	平均
233 動画	1,403,180	10.44

図 4 に、縦軸に「233」の出現回数、横軸に出現回数による順序をプロットしたグラフを示す。図 4 に

示すように、233 コメントの分布は冪分布(power law)に従っている。

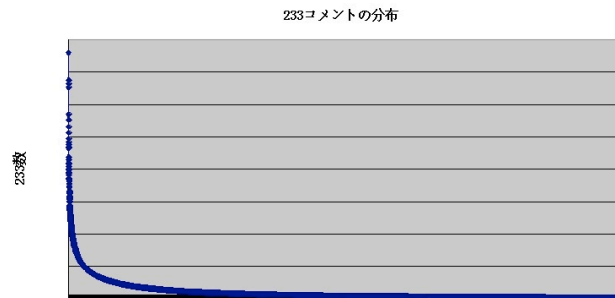


図 4 全 233 コメントの分布図

4 相関分析と考察

4.1 人気の度合い

コメントに 233 が多い動画は、笑える動画であろうと。大笑いできる動画は、人気の高い動画であると考えられる。その実証のため、人気の度合いと、233 コメント数についての相関を調べる。

人気の度合いに使える数値に、再生数(放送数)、長コメント数、弹幕コメント数、マイリスト数(ブックマーク者数)、評点、コイン数がある。これらの値は多いほど人気があると言える。

評点とコイン数はBilibiliが提供するユニークな視聴者評価システムである。この二つの数値は登録会員だけが与えることができる。登録会員が与えられる評点およびコインは、視聴時間および接続時間に比例する。ログインすると 1 日に 40 評点が貰え、また視聴時間が 1 時間につき 0.1 コイン貰える。1 人の会員は、シナリオ動画に最大 2 個コインを付与できる。

再生数は、会員でない人の視聴や、機械アクセスでも増える。一方、評点およびコイン数は、会員でないと与えられず、かつ利用度合いに比例した値しか付与できない。このため評点およびコイン数は視聴者から動画に対する支援度合いを表すと言える。

4.2 233 コメントと各人気度合いの相関

233 コメントの多い動画が、人気の高い動画であることを確認するため、各人気度合いと 233 コメントの相関を計算した。以下の表 5~8 に相関係数を示す。表では、P:再生数、N:長コメント数、D:弹幕コメント数、M:マイリスト数、H:評点、C:コイン数、233:233 コメント数である。

表 5 を見ると、再生数と評点の相関が高いことが分かる。マイリスト数と、評点・コイン数の相関も

高い。一方、コインと評点の相関は 0.56 で低くはないものの、高くもない。233 コメント数は、弹幕コメント数との相関は高いものの、他の数値との相関は少ない。233 コメント多いからといって人気があるとは言えない。

表 5 全 233 動画での相関係数

	P	N	D	M	H	C
N	0.38					
D	0.35	0.22				
M	0.61	0.37	0.26			
H	0.78	0.43	0.30	0.70		
C	0.47	0.51	0.21	0.70	0.56	
233	0.31	0.16	0.56	0.18	0.23	0.19

我々は、コイン数が最も人気度合いを反映すると考えている。そこで、コイン数の多い順に、上位 100, 1000, 10000 個の動画に限定した相関も算出した。その結果を表 6~8 に示す。

表 6 コイン順 top10000 動画 233 コメントの相関係数

	P	N	D	M	H	C
N	0.28					
D	0.12	0.11				
M	0.50	0.27	0.05			
H	0.70	0.34	0.04	0.61		
C	0.37	0.45	0.10	0.63	0.47	
233	0.15	0.06	0.50	-0.02	0.03	0.06

表 7 コイン順 top1000 動画 233 コメントの相関係数

	P	N	D	M	H	C
N	0.17					
D	-0.05	0.03				
M	0.31	0.16	-0.13			
H	0.61	0.30	-0.17	0.49		
C	0.21	0.45	0.01	0.47	0.34	
233	0.12	-0.003	0.47	-0.17	-0.05	-0.03

表 8 コイン順 top100 動画 233 コメントの相関係数

	P	N	D	M	H	C
N	0.23					
D	-0.17	-0.04				
M	0.60	0.14	-0.31			
H	0.70	0.51	-0.35	0.71		
C	0.33	0.54	0.01	0.41	0.53	
233	-0.06	-0.06	0.53	-0.23	-0.22	-0.12

やはり再生数と評点の相関が高い。またマイリスト数と、評点・コイン数の相関は高い。233 コメント数は、弹幕コメント数との相関は高いものの、他の数値との相関は少ない。こちらも、233 コメント多いからといって人気があるとは言えない。

5. おわりに

動画サイトや商品サイトにおける、利用者からのコメントは、その対象についての感情や意見を含む。我々は利用者が付与するコメントに基づく検索や推薦を検討している。本論文では、中国の bilibili 動画を対象に、視聴者コメントの感情を分析した。

感情分析の第一歩として、笑いの感情を表す「233」について分析した。各種の人気度合いと、233 コメントの数について相関を算出した所、233 コメント数は、弹幕コメント数との相関は高いものの、他の数値との相関は少ないため、233 コメント多いからといって人気があるとは言えないことがわかった。

本稿は第一段階の分析結果報告であり、今後多くの詳細分析が必要である。233 を含む様々な感情コメントを用いたランキング手法を提案し、それについて有効性を検証していきたい。

謝辞

本研究はJSPS 科研費 2350099 の助成を受けたものである。

参考文献

- [1] Naomichi Murakami, Eisuke Ito, Emotional video ranking based on user comments, Proc. of iiWAS2011, ACM, pp.499-502, 2011.
- [2] 村上直至, 伊東栄典: 動画コンテンツの視聴者コメントに基づくランキングとその評価, 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), 日本データベース学会, F8-3, 2012.
- [3] Bilibili (Jan.20,2014). In Wikipedia: The Free Encyclopedia. Retrieved from <http://zh.wikipedia.org/wiki/Bilibili>
- [4] 哔哩哔哩 (Jan.20,2014). In Wikipedia: The Free Encyclopedia. Retrieved from <http://ja.wikipedia.org/wiki/%E5%97%B6%E5%93%A9%E5%97%B6%E5%93%A9>
- [5] 櫻井孝昌: 日本が好きすぎる中国人女子, PHP 新書, ISBN-10: 4569812422, 2013.