九州大学学術情報リポジトリ
Kyushu University Institutional Repository

動画サービスにおける視聴者コメントの分析

村上, 直至
九州大学大学院システム情報科学研究府 | 九州大学情報基盤研究開発センター

伊東, 栄典
九州大学大学院システム情報科学研究府 | 九州大学情報基盤研究開発センター

https://hdl.handle.net/2324/19365

出版情報：火の国情報シンポジウム2011, pp.B-4-5-, 2011-03-18. 情報処理学会九州支部
バージョン：
権利関係：

動画サービスにおける視聴者コメントの分析

村上直至† 伊東栄典‡

あらまし 近年，YouTube や Dailymotion に代表される動画共有サービスが人気を呼んでいる． 我々はそのひとつであるニコニコ動画に着目し，動画の推薦について研究している．本稿では，視聴者が動画に付与したコメントを 2-gram に分解し，TF-IDF 値を用いて動画毎の特徴を捉えた．
キーワード 動画共有サービス，folksonomy，ニコニコ動画

Analysis of the comments for video hosting service

Naomichi Murakami† and Eisuke Ito‡

Abstract A lot of video hosting services are provided on the web such as YouTube, and Nicovideo. Nicovideo is a poplar movie uploading and streaming service in Japan. Videos are given not only metadata such as title, creator, upload date and video length by the creator, but also a lot of annotation such as comments by viewers. In this paper, we decompose comments by 2-gram , captured some features of videos using the tf-idf weight.
Keyword video hosting service, folksonomy, Nicovideo

† 九州大学大学院システム情報科学研究院，福岡市
Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka-shi, 812-8581 Japan
‡ 九州大学情報基盤研究開発センター，福岡市
Research Institute for Information Technology, Kyushu University, Fukuoka-shi, 812-8581 Japan
E-mail: † 2IE10084E@s.kyushu-u.ac.jp, ‡ ito.eisuke.523@m.kyushu-u.ac.jp

1. はじめに

近年 YouTube[1]，Dailymotion[2]に代表される動画共有サービスが流行している．これらのサービスでは毎日大量の動画が投稿されており，膨大な数の動画から利用者が求める面白い動画を発見することは難しい．

我々は国内で人気の高いニコニコ動画[3]を対象に動画の推薦・分類に関する研究を行っている[4,5,6]．ニコニコ動画は 2006 年 12 月 12 日に運営を開始した利用者投稿型の動画共有サービスである．運営開始から 4 年以上が経過した 2011 年 1 月現在 540 万件以上の動画が投稿されている．ニコニコ動画には独自のコメントシステムが存在する．(YouTube 等にも，性質は異なるもののコメントシステムは存在する．) 動画の内容に則したコメントは，動画の分類に利用できる．本稿ではニコニコ動画のコメント機能に着目した動画の分類手法を提案する．

2. 視聴者コメントの分析

2.1 視聴者コメント

動画を分類するにあたり，視聴者が付与する情報は Folksonomy な資源として有用である．ニコニコ動画において，動画に対し視聴者が付与できる情報は次の 6 つがある：動画タグ，視聴者コメント，マイリスト登録，動画再生数，コメント数，マイリスト数．本稿ではこのうち，視聴者コメントについて分析を行っている．

視聴者コメントとは，視聴者が動画の再生画面上にコメントを付与することができる機能である．付与されたコメントはサーバで保存され，付与対象となった動画と共に全ての利用者から共有・閲覧される．全く別の時間，別の場所で動画を視聴した，様々な利用者の反応を動画とセットで楽しむことができ，人気を呼んでいる．付与されたタグは 3 秒間かけて，再生画面上を右から左へ横切るように流れる．コメント付与時のコマンドによっては画面の上部，下部等に表示させることもでき，その表示時間は全て 3 秒である．動画の内容に関連したコメントも多く，これは動画の分類に利用できる可能性がある．

2.2 分析手法

ニコニコ動画において 2009 年 11 月時点で「音楽」タグの付与されていた全ての動画 385,280 件に対し，コメント情報の取得を行った．コメントは 2010 年の 5 月から 6 月にかけて，それぞれの動画で直近の 500 件ずつを取得した．付与されたコメント数が 500 件に満たない動画については，その全てを収集している．

視聴者コメントは口語的であり，意図された誤植やインターネットミームを多数含む．そのため ChaSen[7]や MeCab[8]といった辞書型の形態素解析器による解析には不

ⓒ2009 Information Processing Society of Japan

向きである．本論文ではコメント表現を 2-gram に分解して解析を行った．

また，分野に特有な表現を抽出するために TF-IDF 値を調べた．全ての 2-gram 表現に対し，DF ファイルを収集した全てのコメント情報を統合して生成した．TF ファイルは動画毎に生成したものと，タグ毎に生成したものの二種類を用意し，分析にあたった．

2.3 動画毎の特有表現

表はコメント文から 2-gram で抽出した表現を動画毎に，TF-IDF 値の降順で並べたものである．2-gram 表現は空白文字を含むためダブルクォートで囲った．2-gram 表現の隣接右マスには算出した TF-IDF 値を記した．動画タイトルは以下の通りである．

sm1097445:”【初音ミク】みくみくにしてあげる♪【してやんよ】”
sm2726346:”100 分間耐久 【トゥルトゥルダダダ】ベリーナイスジャンキー組合”
sm4187243:”QUEEN-WE WILL ROCK YOU”

表 1：動画毎の特有表現

	sm1097445		sm2726346		sm4187243	
1	く❀	0.0079	”dd”	0.022	“e “	0.0022
2	☆☆	0.0073	“ｄ ｄ”	0.016	“ w”	0.0022
3	❀に	0.0067	“゛ D”	0.012	“ou”	0.0018
4	☆み	0.0049	“A ッ”	0.0080	“ y”	0.0015
5	みっ	0.0047	“DA”	0.0079	“l “	0.0015
6	っく	0.0040	“AD”	0.0052	“wi”	0.0015
7	❀み	0.0040	“ ਾਦ ”	0.0044	“ll”	0.0015
8	く☆	0.0038	“ ਰਮ ”	0.0040	“yo”	0.0015
9	♪⋆	0.0033	“ ਦਲ ”	0.0040	“il”	0.0014
10	⋆♪	0.0033	“ ਲੇ ”	0.0040	“ck”	0.00089

sm1097445 は音声合成ツール「初音ミク」ブームの火付け役となった動画の一つである．動画を彩るために記号を多用したコメントが多く見られた．sm2726346 はインドの楽曲「Tunak Tunak Tun」を 100 分間リピートし続ける動画で，笑いの要素が強い．コメントには”ヾ(゜∀゜)ﾉ゛DADADAッ！”という表現がよく用いられており，2-gram で分析した際にも上位に出現した．また，ヒンディー語に用いられるインド系文字も多用され，混沌としている．sm4187243 はイギリスのロックバンド「QUEEN」の楽曲「We will rock you」のプロモーションビデオである．歌詞”we will rock you”の一部が上位に出現した．

2.4 関連研究

ニコニコ動画のコメント機能は動画の盛り上がりを演出する重要な役割を担っている．中村ら[9]は一度利用者が視聴した動画についてコメントの集中する箇所をサムネイル化して保存，再度その動画を閲覧する際に素早く動画の見所まで移動するシステムを提案している．

ニコニコ動画は仕様変更や投稿動画のトレンド等の変化が激しい．国内での人気の高さから，自発的に研究・解析を行ってきた利用者も多く，それらはニコニコ動画サービスの変化に柔軟である．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[10]を開催した．末永匡は，榊剛史らの論文[11]に倣い，χ^2 値を用いた単語間の関連度の指標に従って動画タグ間の関連度を求めている[12].

動画分類の手法としては，視聴者コメントを用いる以外にも，投稿者間のつながりを用いる方法や，動画タグを用いる方法がある．濱崎らはニコニコ動画における初音ミク動画コミュニティを対象として動画作者ネットワークを分析した[13]. また，我々も動画タグの共起頻度を利用した動画検索支援システムを構築している[6].

3. おわりに

本稿では，動画に付与される 6 つの Folksonomy な情報のうち視聴者コメントのみに着目している．しかし，動画タグはより直接に動画分類に寄与している情報である．過去に我々は動画タグの親子関係を抽出し，動画検索支援システムを作っている．今後はタグとコメントとの関係を活かした検索支援システムを構築し，システムの持つ serendipity 性を向上させたい．

また，今回の分析では 2-gram での表現切り出しに止まったが，N-gram に拡張した解析を行うことで，より意味の分かりやすい特徴表現の抽出を行いたい．

ⓒ2009 Information Processing Society of Japan

参考文献

1) YouTube, http://www.youtube.com, (accessed 2011.2.9).
2) Dailymotion, http://www.dailymotion.com, (accessed 2011.2.9).
3) ニコニコ動画, http://www.nicovideo.jp, (accessed 2011.2.9).
4) 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.
5) 村上直至, 伊東栄典: 共感度検索のための動画アノテーション分析, FIT2010(第 9 回情報科学技術フォーラム), F-002, 第 2 分冊, pp.343-348, Sep.,2010.
6) 村上直至, 伊東栄典: 動画投稿サイトで付与された動画タグの階層化, MPS2010(第 81 回 MPS 研究会), B2-18,Nov., 2010.
7) ChaSen—形態素解析器,http://chasen-legacy.sourceforge.jp, (accessed 2011.2.9)
8) MeCab ホームページ,http://mecab.sourceforge.jp, (accessed 2011.2.9)
9) 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, human interface, 10(4), 103-110, Nov., 2008.
10) 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 2010.10.15)
11) 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, Apr., 2007.
12) 末永匡: χ2 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 2010.10.15)
13) 濱崎雅弘, 武田英明, 西村拓一: 動画共有サイトにおける大規模な協調的創造活動 の創発のネットワーク分析 -ニコニコ動画における初音ミク動画コミュニティを対象と して-, 人工知能学会論文誌, Vol. 25, No. 1, pp.157--167, 2010..

ⓒ2009 Information Processing Society of Japan