

文書分類手法による炎上動画検出手法の検討

竹内, 幹太
九州大学工学部電気情報工学科

伊東, 栄典
九州大学情報基盤研究開発センター

<https://hdl.handle.net/2324/4740663>

出版情報 : 情報処理学会研究報告. 2021 (B3-3), pp.1-4, 2021-03-01. 情報処理学会九州支部
バージョン :
権利関係 : ©2021 Information Processing Society of Japan

文書分類手法による炎上動画検出手法の検討

竹内 幹太^{1,a)} 伊東 栄典^{2,b)}

概要: 動画投稿サイトのコメントの炎上や, SNS での応答が炎上し, 誹謗中傷に発展する事がある。最近では誹謗中傷を苦にして自殺に至る人がでており, 炎上は大きな問題になっている。本研究ではニコニコ動画を対象に, 動画コメントが炎上している動画の検出手法を検討する。具体的には, 動画コメントの単語頻度, 単語の感情度, コメント全体の感情を数値化する。数値化したデータに基づき, 機械学習を援用した文書分類手法を用いて, コメントが炎上している可能性のある動画の選出を目指す。本論文では, 対象データと提案する炎上動画抽出手法について述べる。

キーワード: 動画共有サービス, 利用者コメント, 炎上, 文書分類

KANTA TAKEUCHI^{1,a)} EISUKE ITO^{2,b)}

Abstract: Blowing up of comments on SNS (Social Network System) or video sharing service is a serious problem. Recently, there are two news about suicide of a YouTuber and a TV show cast. Both of them were suffering from slanderous comments. In this paper, we will study a method to detect a blowing up comment in video sharing service, nicovideo.jp. We focus on only comments posted by audiences. To vectorize many comments, we use word frequency, sentiment level of words, and the emotions of the entire comment. Based on the quantified vector and using a document classification method, we try to classify videos whose comments may be blowing by users or not. In this paper, we describe the target data and the proposed flaming video extraction method.

Keywords: Video sharing service, user comments blowing up, document classification

1. はじめに

動画投稿サイトのコメントの炎上や, SNS での応答が炎上し, 誹謗中傷に発展する事がある。最近では誹謗中傷を苦にして自殺に至る人がでており, 炎上は大きな問題になっている。動画サービスでも, 視聴者が投稿するコメントが荒れ, 誹謗合戦になることもある。このように現代社会では誹謗中傷コメントの発見や, 炎上案件の発見は大きな意味がある。

本研究ではニコニコ動画における炎上動画の検出を検討する。ニコニコ動画はニコニコデータセット [1] として動画のメタデータ・コメントデータが提供されている。ニコニコ動画では独自のコメント機構を備えている。そのため 1 つの動画に多数のコメントが付与されやすい。また, 炎上した動画についての情報も提供されている。付属サービ

スのニコニコ大百科は Wikipedia に似ており, 多数のエントリに関する記事が作成されている。記事の 1 つに炎上動画に関する記事があり, それらの動画群を機械学習の正例として利用できる。またニコニコ動画独自のタグ付けシステムも利点がある。視聴者が付与するタグを用いて炎上の可能性が高い動画を検索できる。

本論文では教師あり機械学習による文書分類機を用いて, 誹謗中傷コメントで荒れた炎上動画の抽出を検討する。教師あり機械学習の性能は, 学習用の正解データ, 分類対象のベクトル化, および分類アルゴリズムに依存する。炎上動画の検出には, 学習用の正例 (炎上) 動画を集めなければならない。コメントが誹謗中傷で荒れている動画を炎上動画とするため, 動画の数値ベクトル化として動画コメントを利用する。単語頻度, 単語の感情度, 各コメントの感情度などを検討する。

本論文の構成を述べる。第 2 節で本研究で使用したニコニコデータセットを説明する。第 3 節では, 学習用データの選定, 動画コメントのベクトル化についての検討結果を述

¹ 九州大学工学部電気情報工学科

² 九州大学情報基盤研究開発センター

a) takeuchi.kanta.881@s.kyushu-u.ac.jp

b) ito.eisuke.523@m.kyushu-u.ac.jp

べる。第4節では、小規模データでの実験結果を述べる。最後に第5節でまとめと今後の課題を述べる。

2. 対象データ

2.1 ニコニコ動画

ニコニコ動画^{*1}は、2006年12月開設の日本の代表的動画共有サイトで、ドワンゴ社の運営である。付属するニコニコ大百科^{*2}は大百科ニュース社の運営である。カドカワ社の2017年3月期通期決算説明資料[9]によれば、2017年3月において、会員数6430万人、うち有料会員数243万人である。

利用者はマイリストにお気に入り動画を登録できる。利用者は閲覧動画にコメントやタグを投稿できるし、自分の独自動画を投稿できる。利用者のうち、多くの利用者は動画を閲覧するだけである。一部の利用者がコメント投稿やタグ付与をしている。動画の投稿者は、全利用者数に比べると極一部である。

ニコニコ動画のコメント機能は独自である。動画の再生時間に対応してコメントは投稿される。動画再生時には画面上にオーバーレイする形でコメント文が流れて表示される。またCGM百科事典の「ニコニコ大百科」でタグの説明記事を作成できる。これらの独特なシステムにより、ニコニコ動画では、他の動画共有サイトには無い、独自の文化を築いて来た。各動画が細かなカテゴリに分類されているのもニコニコ動画の大きな特徴の一つである。

2.2 ニコニコデータセット

ニコニコデータセットはドワンゴ社および大百科ニュース社から提供されたデータで、国立情報学研究所が研究用に公開している。ニコニコ動画のサービス開始当初から2018年11月8日までの投稿された約1,670万件の動画のメタデータと、それらの動画に対する約38億件のコメントデータを含む。動画のメタデータはJSON形式で格納されている。各コメントデータは、1つの動画のコメントが1ファイルにJSONL形式でまとめられている。

表1に、ニコニコデータセットのデータ取得期間やデータ件数などを示す。

表1 ニコニコデータセット概要

	詳細
対象期間	2007年3月6日 - 2018年11月8日
総動画数	16,703,325
総コメント数	3,773,083,461

*1 <https://nicovideo.jp>

*2 <https://dic.nicovideo.jp>

表2 動画メタデータの属性

属性名	内容
video_id	動画ID
title	動画タイトル
description	動画説明文
watch_num	再生数
comment_num	コメント数
mylist_num	マイリスト登録数
category	動画カテゴリ
tags	タグ
upload_time	投稿日時
file_type	動画フォーマット
length	再生時間
size_high	高画質動画のファイルサイズ
size_low	低画質動画のファイルサイズ

表3 コメントデータの概要

属性名	説明
date	コメント投稿日時
vpos	縦方向の位置
content	コメント文
command	各種命令

2.3 動画投稿数など

図1にニコニコ動画の月別動画投稿数の推移を示す。2015年頃まで動画投稿数は増加傾向にあり、2015年3月に約1万2千件となりピークに達した。その後2016年7月を境に動画投稿数が減少しており、2018年3月の投稿動画数は約7千件である。

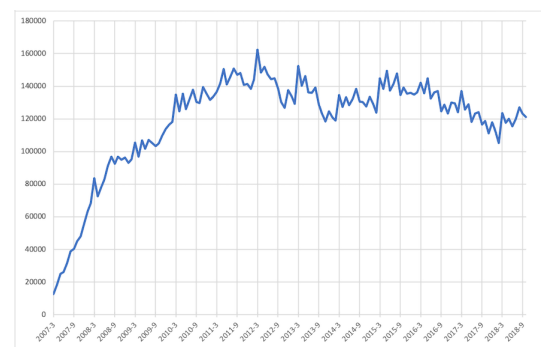


図1 月毎の動画投稿数

3. 炎上動画の抽出手法

炎上動画か否かの判定は、基本的な2クラス分類問題になる。教師あり機械学習で分類する場合、正例・負例の訓練データの収集、対象の数値ベクトル化手法、機械学習による分類手法、の3つが問題になる。

3.1 学習用の炎上動画選定

教師あり機械学習では、正例と負例の学習用データが多

数必要である。炎上動画を正例、そうでない動画を負例とする。膨大な数の動画で、殆どのコメントは炎上していない。正例となる炎上動画はごく少数であるため、正例の選定は難しい。

最初にニコニコ大百科の「炎上」記事 [2] からの炎上動画選定を検討した。記事内には過去に炎上した動画の説明とリンクがある。しかしながら、これらの炎上動画は既に削除されている。削除動画の動画メタデータおよびコメントは、データセットで提供されていない。

次に動画サイトの動画評価機構の利用を検討した。ニコニコ動画には視聴者が「いいね」と「ニコる」を投票する仕組みがあり、データセットには投票数が格納されている。しかしながら、この投票数は炎上動画選定に使用できない。理由は、比較的最近の動画にのみ投票数が在ることと、そもそも投票は良い動画を推すもので悪い炎上動画選定に使用できないためである。

最終的に、タグによる炎上動画候補と人力による判断で炎上動画を選定した。炎上動画と思われる動画を 100 個程度見たところ、特定タグが付与されていることが分かった。表 4 に炎上動画の候補選定に用いたタグを示す。表 4 に示す単語をタグに 1 つでも含み、再生数が 1000 以上、コメント数が 1 万以上の動画を候補とした。候補の動画を人手で見て、コメントが荒れている炎上動画を 848 個選定した。

表 4 炎上動画候補に用いたタグ

う p 主は知的障害者シリーズ, 人間のクズ, この顔にピンときたら 1 1 0 番, 社会のゴミ, 老害, 統合失調症, 死ぬばいいのに, アンチ, 粘着

最後に負例である「炎上していない動画」を、正例と同数選定した。正例動画を 1 つ選び、その動画の再生回数・コメント数・ジャンル(タグ)を調べる。ジャンル(タグ)が同じで、再生回数およびコメント数が同程度の動画を検索する。そのうち、炎上していない通常の動画を

3.2 動画コメントのベクトル化

本研究では再生される動画データ本体は扱わず、動画に付随するテキストデータを扱う。動画に付随するテキストデータには、動画メタデータ(投稿日・タイトル・作者・タグなど)と動画コメントデータの 2 つが在る。炎上動画ではコメントが荒れているため、コメントから数値ベクトルを作る。本研究で検討したベクトル化手法の概要を図 2 に示す。

文書の数値ベクトル化で、古くから使われる手法は単語ごとの出現頻度である [3]。文書内の文を形態素解析機で単語に分割し、単語の出現回数を調べる。文書に出現する単語の頻度を数え上げれば、文書を数値ベクトルで表現できる。単語頻度を用いる場合、単語の切り分け手法の検討

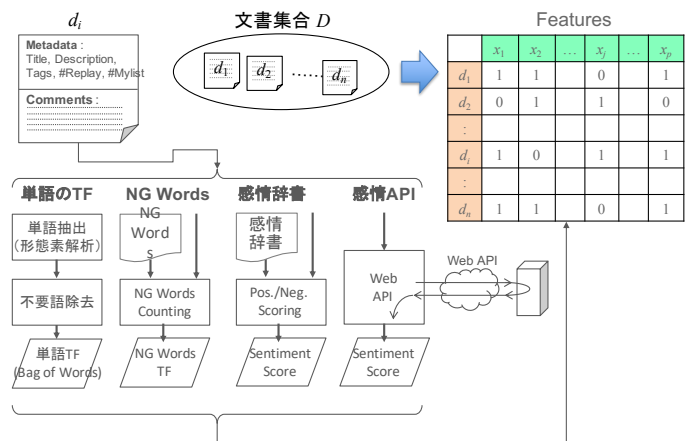


図 2 動画コメントのベクトル化

が必要である。ニコニコ動画における多くの動画コメントは 1 単語あるいは顔文字のような辞書に記載のないものが多いため、一般の形態素解析が適さないからである。

単語頻度を数え上げる場合、NG ワードの利用も検討できる。ニコニコ大百科には「NG 推奨ワード」の記事がある [4]。この記事には、視聴者が不愉快になりがちなコメント単語が記載されている。NG ワードの出現頻度を用いると、コメント炎上か否かの判定に使える可能性がある。

文書のセンチメント分析も援用できる。高村は自身の Web サイトで感情辞書を公開している [5][6]。この辞書では、感情的な単語について、-1 から 1 の範囲で感情数値を付与している。ネガティブが -1 でポジティブが 1 になる。感情辞書に記載のある単語をコメントから拾い上げ、感情辞書のスコアでコメントのポジティブ度・ネガティブ度をスコアにする。感情辞書の援用には 2 つの問題がある。1 つは日本語の感情辞書に登録された単語はネガティブ単語が多いことで、もう 2 つ目の問題はポジティブ・ネガティブの値しかないことである。

感情算出には、ニコニコ動画コメント分析室 [7] の「一言コメント感情分析 API」も利用できる。この分析室は、コメントの感情を 19 種類定義している。ニコニコ動画のコメントを API で送付すると、19 個うちの 1 つを返す。表 5 に 19 個の感情を示す。

次にニコニコ大百科の「NG 推奨ワード一覧 [4]」を手がかりに炎上動画選出を試みた。ニコニコ動画では、視聴者が閲覧拒否したいコメントを「NG ワード」に指定できる。動画コメントに NG ワードを多く含むものは炎上している可能性が高い。そこでコメントに「NG ワード」を多く含む動画を調べ、実際に炎上していると判断した動画を選出した。先の 25 個と合わせて、合計で 1000 件の動画を炎上動画とした。

3.3 機械学習による分類手法

教師あり機械学習で分類問題を解く手法にはいくつかの

表 5 「一言コメント感情分析 API」の返り値

No	返り値	内容	例
1	unknown	分類不能	
2	aa	Ascii Art	顔文字など
3	feed	ツッコミ	「いい加減にしる」など
4	irritate	イライラ	「やれ, しろ, やめろ」など
5	lol	笑い	wの繰返しや「~で草」など
6	like	称賛	良い評価コメント
7	critic	批判	悪い評価やダメ出し
8	cry	感動	「泣いた, 悲しい」
9	request	要望	「~して欲しい, はよ」
10	response	返事	動画投稿者への返事
11	advice	助言	「~した方がいい」など
12	surprise	驚愕	「!」を含むものなど
13	abusive	暴言	「氏ね」など
14	greeting	挨拶	「うぼつ, 乙」
15	exciting	期待	「w k t k, 楽しみ」
16	fever	熱狂	「きゃあああ」や語尾が!
17	question	疑問	「~かな」や語尾が?
18	fear	恐怖	悲鳴や怯え
19	exclude	除外	分析対象外コメント

ものがある [8]。本研究では Python 言語用の scikit-learn モジュール [9] に付随する分類器の利用を検討する。具体的には, SVM (Support Vector Machine), 決定木, MLP (Multi Layer Perceptron) で学習を行い, 出来たモデルを炎上動画分類器とする。

4. 実験と考察

第 3.1 節で述べた正例の炎上動画と負例の炎上でない動画について, 第 3.2 節で述べた動画コメントを用いたベクトル化を適用した。そのデータを用いた機械学習, SVM, 決定木 (D-tree), MLP で分類器を作成した。分類性能の評価尺度は適合率 (Precision), 再現率 (Recall), 正解率 (Accuracy), F 値 (F-measure) を用いる。表 6 に 5 回の 5 分割交差検定の評価値を記載する。なお, 動画コメントをベクトル化する際, API: 一言コメント感情分析 API のみを用いる場合, NG: NG ワードの頻度を用いる場合, API+NG: 両方を用いる場合, の 3 つを比較した。

表 6 を見ると, どの場合も数値が 0.8 を越えており, 学習用データのみでの分類性能は良い。F 値で比較すると, 動画コメントのベクトル化については, NG:NG ワードの頻度のみを用いる場合の, SVM 分類器が最も高い値を示した。ただし, 学習および検証に用いた正例 (炎上)848 件, 負例 848 件のデータ集合に対しての分類性能である。

表 1 に示す様に, ニコニコデータセットには 1600 万以上の動画に対応するデータを保持する。学習用以外の動画データは膨大である。今後は学習用以外の動画のコメントデータ用いて炎上動画の抽出を試し, 分類器の性能を評価

する必要がある。

5. おわりに

本研究ではニコニコ動画を対象に, 教師あり機械学習である文書分類器を用いて, 誹謗中傷コメントで炎上している動画を検出する手法について検討した。動画コメントのベクトル化では, 単語頻度, 単語の感情度, コメント感情度算出 API, NG ワードの出現頻度を用いる手法を提案した。学習用の正例 (炎上) 動画は, 炎上動画に頻出するタグを用いての候補から, 人手判別で選出した。負例は, 正例の動画と再生回数・コメント数・タグの種類が同程度で, 炎上していない動画を選出した。それらの学習用データを用いて, SVM, 決定木, MLP を用いて分類器を作成した。学習用データだけからは, 高い分類性能を実現した。今後は学習用以外の動画のコメントデータ用いて炎上動画の抽出を試し, 分類器の性能を評価する。また, 他のベクトル化手法の検討や, 学習データの増加も行いたい。

参考文献

- [1] 国立情報学研究所: 情報学研究データリポジトリニコニコデータセット, <https://www.nii.ac.jp/dsc/ldr/nico/>.
- [2] ニコニコ大百科: 炎上, <https://dic.nicovideo.jp/a/%E7%82%8E%E4%B8%8A>.
- [3] 北研二, 津田和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版 (2002).
- [4] ニコニコ大百科: NG 推奨ワードの一覧, <https://dic.nicovideo.jp/a/ng/%E6%8E%A8%E5%A5%A8%E3%83%AF%E3%83%BC%E3%83%89%E3%81%AE%E4%B8%80%E8%A6%A7>.
- [5] Takamura, H., Inui, T. and Okumura, M.: Extracting Semantic Orientations of Words using Spin Model, *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005)*, pp. 133-140 (2005).
- [6] 高村大也: 単語感情極性対応表, http://www.lr.pititech.ac.jp/~takamura/pndic_ja.html.
- [7] ニコニココメント感情分析質: 一言コメント感情分析 API, <http://nicomment.ktty1220.me/api.html>.
- [8] 秋庭伸也, 杉山阿聖, 寺田学, 加藤公一: 見て試してわかる機械学習アルゴリズムの仕組み機械学習図鑑, 翔泳社 (2019).
- [9] Scikit-Learn: Scikit-learn Machine Learning in Python, <https://scikit-learn.org/stable/>.

表 6 学習器の分類性能

データ	分類器	Precision	Recall	Accuracy	F-measure
API	SVM	0.8577	0.9080	0.9003	0.8819
	D-tree	0.8207	0.8881	0.8742	0.8527
	MLP	0.8409	0.9081	0.8912	0.8727
NG	SVM	0.9077	0.9175	0.9274	0.9122
	D-tree	0.8161	0.8043	0.8452	0.8097
	MLP	0.8850	0.8857	0.9057	0.8848
API+NG	SVM	0.9069	0.9022	0.9216	0.9043
	D-tree	0.8330	0.8762	0.8766	0.8536
	MLP	0.8900	0.8856	0.9076	0.8873