

特集論文 「Web インテリジェンスとインタラクションの新展開」

CGM における炎上の分析とその応用

Analysis of Flaming and Its Applications in CGM

岩崎 祐貴
Yuki Iwasaki

電気通信大学大学院情報システム学研究所
Graduate School of Information Systems, University of Electro-Communications
y-iwasaki@ohsuga.is.uec.ac.jp

折原 良平
Ryohei Orihara

(同 上)
ryohei.orihara@toshiba.co.jp

清 雄一
Yuichi Sei

(同 上)
sei@is.uec.ac.jp

中川 博之
Hiroyuki Nakagawa

大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University
nakagawa@ist.osaka-u.ac.jp

田原 康之
Yasuyuki Tahara

電気通信大学大学院情報システム学研究所
Graduate School of Information Systems, University of Electro-Communications
tahara@is.uec.ac.jp

大須賀 昭彦
Akihiko Ohsuga

(同 上)
ohsuga@is.uec.ac.jp

keywords: flaming, microblogging, reputation mining, decision tree, sentiment analysis

Summary

Nowadays, anybody can easily express their opinion publicly through Consumer Generated Media. Because of this, a phenomenon of flooding criticism on the Internet, called flaming, frequently occurs. Although there are strong demands for flaming management, a service to reduce damage caused by a flaming after one occurs, it is very difficult to properly do so in practice. We are trying to keep the flaming from happening. It is necessary to identify the situation and the remark which are likely to cause flaming for our goal. Concretely, we propose methods to identify a potential tweet which will be a likely candidate of a flaming on Twitter, considering public opinion among Twitter users. Among three categories of flamings, our main focus is Struggles between Conflicting Values (SBCV), which is defined as a remark that forces one's own opinion about a topic on others. Forecasting of this type of flamings is potentially desired since most of its victims are celebrities, who need to care one's own social images. We proceed with a working hypothesis: *a SBCV is caused by a gap between the polarity of the remark and that of public opinion*. First, we have visualized the process how a remark gets flamed when its content is far from public opinion, by means of our original parameter *daily polarity (dp)*. Second, we have built a highly accurate flaming prediction model with decision tree learning, using cumulative dp as an attribute along with parameters available from Twitter APIs. The experimental result suggests that the hypothesis is correct.

1. はじめに

1.1 背景

近年、ブログや Twitter の発展に伴い、投稿した自分の発言に対し批判的なコメントが殺到する「炎上」という現象 [田代 08] が頻繁に起こるようになった。それにより、ブログの閉鎖やアカウントの削除、個人情報の特定といった被害に遭うリスクが高まっている。既に発生した炎上を抽出するには、批判的なコメントが殺到する状況を検知すればよい [NAVER 10] が、それでは炎上を未然に防ぐことができない。そこで本研究では、炎上事例を収集し機械学習させることで、未来の炎上予測を行い、炎上を未然に防ぐことを目標とする。この目標を達成するためには、炎上しそうな状況や発言を同定する必

要がある。

以下に、本稿での炎上関連の用語法について整理する。すでに述べたとおり、炎上はある CGM 記事に対し批判的なコメントが殺到する状況を指す。炎上を引き起こすきっかけとなった CGM 記事を炎上元と呼ぶ。過去に発生した炎上の事例を炎上事例と呼び、その特定には炎上元の URL が用いられるが、炎上元は炎上発生後しばしば削除されるため、炎上元の投稿者・投稿時期・内容概要を示して炎上を特定することもよく行われる。

1.2 炎上の類型化

炎上は数種類に類型化できるとされており [小林 11]、実際に 100 件の炎上事例 [NAVER 10] を人手で分類したところ、犯罪自慢・価値観の押し付け (Struggles between

表 1 炎上の内訳 (N=100)

	犯罪自慢	SBCV	暴露
有名人	2	35	2
非有名人	49	6	6
計	51	41	8

Conflicting Values: SBCV)・暴露による炎上が全てを占める結果となった(表 1)。

犯罪自慢とは、自らの犯罪行為を公表することであり、非有名人の炎上が圧倒的に多い「無免許」「飲酒運転」「万引き」「 に爆弾を仕掛けた」等、犯罪自慢特有の単語が目立つという特徴がある。SBCV とは、ある話題について自分の評価を強引に押し通すような発言のことで、自分の評価と異なる意見を持っている人が多い場合に炎上しやすくなる。例えば「猪瀬さんは清廉潔白だ」などがあり、犯罪自慢に比べ、有名人の炎上が多い。暴露とは有名人や所属機関の秘め事を公表する発言である。ホテル従業員が有名人カップルの宿泊デートを暴露した例などがある。有名人は自分の社会的イメージを良好に保つことに腐心せざるを得ないため、有名人の炎上の多い SBCV の予測には潜在的需要があると考えられる。

2. SBCV の実例分析と仮説の提示

2.1 SBCV 事例紹介

§ 1 事例 1: Ashton Kutcher

Twitter の熱狂的な利用者で、800 万人以上の Follower を持つハリウッド俳優 Ashton Kutcher が、Pennsylvania 州立大学フットボールチームの名コーチであった Joe Paterno の解雇に対し、「Joe Paterno を解雇だなんてどういうことだ?この目利きの男のファンとして、遺憾に思う」と Tweet したところ、激しい非難が寄せられた。Paterno には同僚の性的虐待を隠匿した疑いが科せられていたが、Kutcher はそれを知らなかった[鯨岡 11]。スキャンダルにより Paterno への世間の評価(世評)が否定的な中、肯定的な発言をすることによって生じた SBCV である。

§ 2 事例 2: 谷亮子

元柔道家で現参議院議員の谷亮子氏が、女子柔道の体罰・パワハラ問題で告発された園田隆二監督に対し「園田監督は素晴らしい人だ」[山口 13] とかばう発言をし、Twitter 上で激しい非難が寄せられた。告発報道で園田監督への世評が否定的な中、肯定的な発言をすることによって生じた SBCV である。

§ 3 事例 3: 宮崎司

某音楽番組で女性アイドルグループである AKB48 と共演したバイオリニスト、宮崎司氏が、その出演を知らせる投稿で、「生理的に受け付けられない AKB の後ろで弾きます」[LivedoorNEWS 13] と発言した結果、批判が殺到し、Twitter アカウントを削除する騒ぎになった。人気者

である AKB48 への世評が肯定的な中、否定的な発言をすることによって生じた SBCV である。

§ 4 事例 4: 楽しんご

元フジテレビアナウンサーのチノパンこと千野志麻氏が起こした死亡事故に対し、お笑い芸人の楽しんご氏が「チノパン可哀想」と加害者を擁護する発言をしたところ、炎上した[東スポ 13]。事故報道で千野氏への世評が否定的な中、肯定的な発言をすることによって生じた SBCV である。

2.2 動的評判分析

SBCV 事例の観察から、世評を捉えることが炎上予防には重要であることがわかった。そこで、その分析には世評を数値化することが有効であると考えられる。そのため、日次極性(daily polarity: dp)を提案する。これは、日毎の肯定的な Tweet 数と否定的な Tweet 数の差をトピックに関する全 Tweet 数で割った値であり、式(1)によって定義される。トピックとは、発言に含まれている話題のことである。dp は、そのトピックの全体的盛り上がりの中で、当日の Tweet から読み取れる世評の極性の強さを表している。dp に対し、その取得範囲の最大値で割ることで正規化したものが ndp (normarized dp) であり、式(2)で定義される。

$$dp_{I,T} = \frac{P_I(t) - N_I(t)}{\sum_{t \in T} (P_I(t) + N_I(t) + E_I(t))} \quad (1)$$

$dp_{I,T}(t)$: 時区 T 中の時刻 t におけるトピック I の日次極性

$P_I(t), N_I(t), E_I(t)$: 時刻 t におけるトピック I に関する肯定的・否定的・中立 Tweet 数

T: トピック I が話題になっている時区間

$$ndp_{I,T}(t) = \frac{dp_{I,T}(t)}{\max_{t \in T} |dp_{I,T}(t)|} \quad (2)$$

縦軸を ndp と日毎の P/N/Tweet 数、横軸を日毎の時間軸としてグラフ化することで、世評を可視化することができる。2.1(2)~(4)に示した事例の可視化例を図 1~図 3 に示す。ある時点で過去一定期間のデータを取得・分析するのみで、分析結果の経時変化に興味を持たない評判分析(例えば[井上 07])では、トピックに関する調査時点以前の累積的態度に着目することになるが、本手法によればトピックに対する世評の時間的推移を観察できるところに特徴があり、動的評判分析と呼ぶことにする。なお、大きさ 1 日の移動窓を用いてデータ取得し、日次の分析を繰り返すことにより、動的評判分析と同様の効果を得ることができるので、動的評判分析はテキスト評価分析手法[乾 06]として新規なものではないが、ndp による極性の定量化と、極性と Tweet 数を同時に表示する可視化に特徴がある。本論文では、ニュース等の影

響で極性が変化するトピックを動的トピックと呼び、動的トピックのうち、極性が肯定・否定・中立であるものを動的 P/N トピックと呼ぶ。動的評判分析により、世評の変化を朗報や悲報などの影響や時間経過を考慮に入れて分析することが可能となる。なお、世評の時間的推移に着目した研究はこれまでもあるが、それらとの違いについては 7.3 節で触れる。

2.3 SBCV 発生メカニズム仮説

2.1 に示した事例を観察することから、SBCV の発生メカニズムとして、次の仮説を掲げる: あるトピックに対する発言者の評価と世評が対立する場合に炎上が発生する。次節で動的評判分析を用いてこの仮説の妥当性を検討する。

2.4 SBCV 事例の動的評判分析

§1 事例 1: Ashton Kutcher

名門フットボールチームへのコーチとしての功績を考えると、不祥事が発覚する前は、Joe Paterno の評判が肯定的（少なくとも中立以上）であったと仮定できる。しかし不祥事後、このトピックの世評は大きく否定に動いた。Joe Paterno の世評の変化を見抜けなかったことが Kutcher の失敗であり、世評が否定に傾いていることを可視化できれば、発言前に炎上を防ぐことができると考えられる。

この事例における Joe Paterno は動的トピックである。表 1 の SBCV41 例のうち 32 例は、動的トピックに関連しており、その分析の必要性が伺える。

§2 事例 2: 谷亮子

この事例に対し、提案した動的評判分析が過去の炎上事例で有効であるかを確かめる実験を行った。動的評判分析により、世評の時間的変化を分析するため、炎上の起こった日を含む 2013/1/23 から 2013/3/24 までの約 2ヶ月間における園田監督に関連する Tweet を収集した。すなわち、検索クエリ“園田監督 OR 園田隆二”に該当する Tweet 全 14385 件を対象とした。図 1 は日次極性を肯定的/否定的 Tweet 数と共に表したもので、日次極性が肯定的/否定的 Tweet 分布の推移をよく表していることがわかる。

世評の取得範囲の設定について考察する。範囲を広くしすぎると自動極性判定に時間がかかり、可読性も悪くなる。狭くしすぎると同じトピックで異なる盛り上がりや長期的な盛り上がりが起こったときに発見しにくい。そのため本区間を適切に設定する必要があるが、本事例では経験的に 2ヶ月に設定した。

前述した Kutcher の例と同様に、園田監督に対する世評は、体罰問題の報道以前は中立であったと考えられ、その時点ではこの発言は炎上しない。本研究では、これを発言の趣旨が世評と対立しないからであると考えられる。

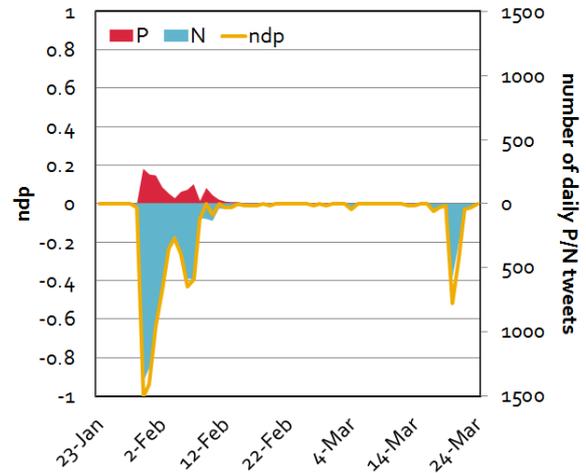


図 1 世評の推移 (トピック: 園田監督)

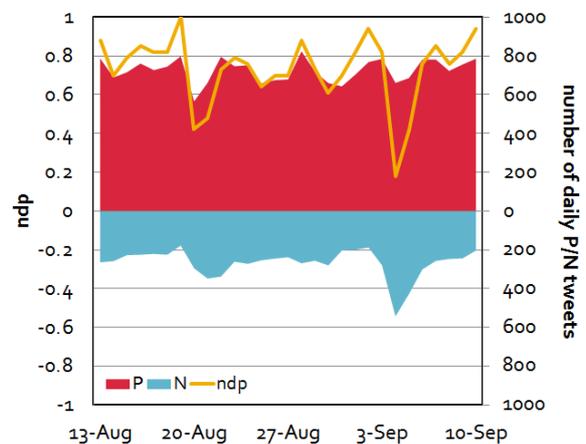


図 2 世評の推移 (トピック: AKB48)

一方、体罰問題の報道以降、園田監督の世評は否定的になったと考えられる。そこに「素晴らしい」という肯定的な、世評からずれた発言をしたことで炎上した、というのが本研究のアプローチに基づく解釈である。

§3 事例 3: 宮崎司

この事例に対し、世評の時間的変化を分析するため、炎上の起こった日を含む 2013/8/13 から 9/10 までの約 2ヶ月間における AKB48 に関連する Tweet を収集した。すなわち、検索クエリ“AKB”に該当する Tweet 全 53178 件を対象とした。トピック: 園田監督とは異なり、全体的に Tweet 量が多く、極性が肯定的に安定していることが伺える (図 2)。

否定的な発言が炎上した 9/3 直前に注目すると、ndp は 0.82 であった。つまり、この時点での AKB48 に対する世評は全体の 8 割以上肯定に傾いていたと考えられ、発言と世評の対立が炎上を招いたと説明できる。なお、9/3 の当炎上発言のニュース化に伴い、炎上同様に批判的なコメントが殺到し、ndp は 0.18 と低下した。また、8/20 時点において ndp が多少否定方面に傾いているのは、別

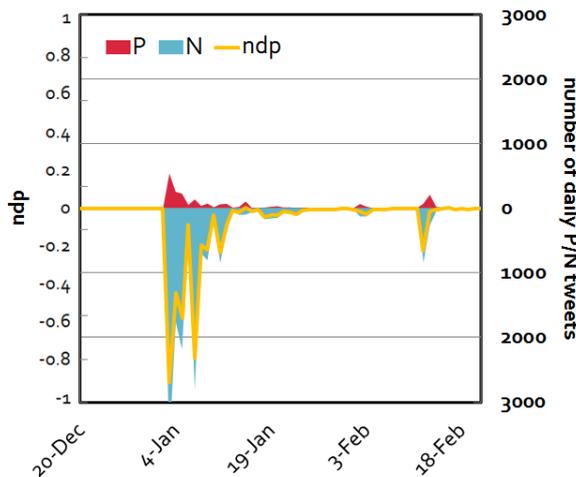


図3 世評の推移 (トピック: チノパン)

の炎上が起こっていた [Togetter 13] ためである。

§4 事例4: 楽しんご

この事例に対し、炎上の起こった日を含む 2012/12/20 から 2013/2/20 までの約 2ヶ月間におけるチノパンに関連する Tweet を収集した。すなわち、検索クエリ“チノパン OR 千野志麻 OR 千野アナ”に該当する Tweet から目視で不適切な Tweet^{*1} を除去した全 24189 件を対象とした。図3はトピック:チノパンの世評の推移を可視化したもので、2013年1~2月の否定的な谷が4つ抽出できていることがわかる。

2013/1/2 から 1/12 に注目すると、否定的な Tweet が劇的に増加していることがわかる。この否定的な Tweet の谷は、死亡事故を起こしたチノパンというトピックに対する世評が否定的であることを反映していると推測できる。炎上した発言「チノパン可哀想」は、チノパンを擁護する肯定的な発言であると言える。つまり、トピック「チノパン」について、楽しんご氏の肯定的な意見と否定的な世評が価値観のずれを起こし、炎上に繋がった事例であると解釈できる。

なお、期間毎の Tweet から抽出した特徴語をニュースと対応付ける方法でも世評を可視化できる [岩崎 13]。例えば、死亡事故直後には「書類送検」、数日後には「通夜」、1週間後には「降板」といった特徴語が抽出され、これらから報道内容の移り変わりを見て取ることができる。

3. 提案手法

2章で分析した事例に現れるトピック中、AKB48 以外は動的トピックである。2.3 に示した作業仮説に従えば、世評が肯定に安定している AKB48 のようなトピックでは、その世評に逆らわないようにしていれば炎上しない。ところが園田監督やチノパンのように世評が変化するト

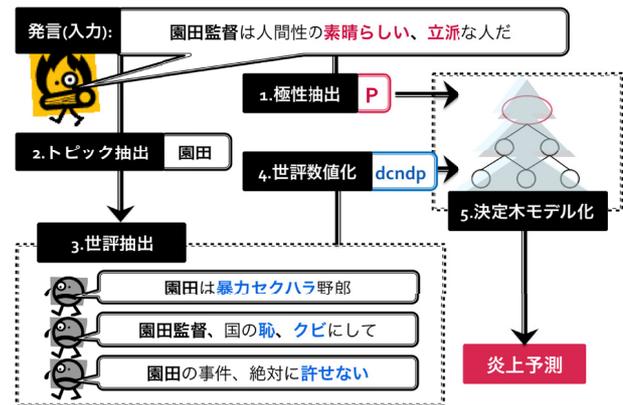


図4 提案手法概要

ピックでは、変化し続ける世評を読み取った上で適切な発言が求められるため、扱いが難しい。そこで、発言内容に動的トピックが含まれる場合でも、炎上を未然に防ぐことを目的とした SBCV 予測手法を提案する (図4)。

なお、CGMにおける炎上元はブログや Twitter など様々なメディアがあり得るが、本研究では、発言や発言者に関する豊富な属性を利用できることから、対象を Twitter に限定する。Twitter における炎上元を炎上 Tweet、炎上 Tweet を含む一定時期の Tweet 集合を炎上データと呼ぶ。

これから Tweet しようとする発言を入力とし、手法の流れを順に説明する。

3.1 発言の極性抽出

炎上元となり得る発言に対し、辞書マッチングによる評判分析を行い、発言の極性を自動的に抽出する。具体的には、入力 Tweet を形態素解析器 [Kudo 04] に通し、用言 5281 語、名詞 13315 語を含む P/N 辞書 [小林 05, 東山 08] とマッチする語のうち、P/N の数の多い方をその Tweet の極性とした。

3.2 発言のトピック抽出

世評を抽出するためには、対象の発言からトピックを抽出する必要がある。トピックの抽出方法として、隣接語の共起関係などが挙げられる [奥村 05]。しかし、本研究での目的はトピック推定ではなく、トピックが正しく推定できた後の炎上分析の精度であることから、本稿では人手でトピックを入力する形とする。

3.3 世評抽出

評判分析器を TwitterAPI [Twitter 13] により収集した Tweet 群に適用することで、世評の極性を自動的に抽出する。計算対象の世評を把握するために時間的な範囲を指定する必要がある。本研究では、炎上 Tweet となることが予想される Tweet が投稿されてからその影響が現れるまでについて、1週間では短く、1ヶ月では長すぎるといった経験則と、炎上後の極性の動きを分析するため、そ

*1 チノクロス製パンツ、通称チノパンなど。

の収集期間を、炎上元となり得る Tweet 日の 3 週間前から 1 週間後までの約 1ヶ月間とした。

3.4 世評数値化

2.3 に示した作業仮説に基づき, ndp を用いてある時点での世評を数値化することを考える。ndp は, ある 1 日の極性の強さしか表現していない。例えば, ニュースによって ndp が変化する場合, ニュース報道直後には ndp が大きな絶対値を取り, その後時間経過に伴って ndp の絶対値は急激に減少するが, ニュースが記憶されている期間は世評の極性の変化は ndp の変化に比べて緩やかであると考えられる。そのため, 発言中の評価と世評のずれを数値化するためには, 過去の ndp の影響を積分した値を用いる必要がある。そこで, 割引累積日次極性 (discounted cumulative ndp: dcndp) を提案する。これは以下の式 (3) で定義され, これと発言の極性を併せて観察することで発言と世評の対立の有無を検出できる。ニュース等による印象が世間から忘れ去られていく様子を表すために, 忘却係数 0.95 を用いている。

$$dcndp_{I,T}(t) = \sum_{i=1}^{\infty} ndp_{I,T}(t-i) * 0.95^{i-1} \quad (3)$$

3.5 決定木モデル化

過去の Twitter データを収集し, Tweet が炎上したかどうかを手で判定し教師データとした上で, Tweet の極性と数値化した世評, 及び TwitterAPI から得られる各種パラメータを属性とし, 炎上の有無を分類する決定木を学習することで SBCV 予測モデルを構築する。

4. データ収集

3.5 に示した通り, 提案手法の実現には多くの炎上 Tweet が必要だが, 炎上事例を収集した Web サイト [NAVER 10] 上の事例は, 炎上 Tweet が削除されているものが多く炎上 Tweet 収集には不向きである。そのため, 削除されていない炎上 Tweet を収集するモデルを提案する。

基本的なアプローチは, 少数の炎上 Tweet を目視で探し出し, それらを非炎上 Tweet と弁別するモデルを決定木により学習し, 同モデルを大量データに適用して炎上 Tweet の候補を精度良く収集するというものである。本研究では, 最初に種とした 37 件の炎上事例に基づき学習したモデルを 1533 件の Tweet に適用したところ, 24 件増やすことに成功した。

4.1 決定木で扱う属性

TwitterAPI から得られる属性の他, それを応用した属性 (表 2) を元に決定木分類を行うことで, 炎上 Tweet 分類器を作成する。なお, 以下の説明では ReTweet を RT, Favorite を Fav と略記する。

表 2 決定木分類器で扱う属性

属性名	説明
Follow 数	Tweet したアカウントの Follow 数。
Follower 数	Tweet したアカウントの Follower 数。
平均 RT 数	Tweet したアカウントの最新 Tweet3200 件から取得した各 RT 数の平均値。最新 3200 件とは, 1 つのアカウントから TwitterAPI を使って取得できる Tweet の上限である。
平均 Fav 数	Tweet したアカウントの最新 Tweet3200 件から取得した各 Fav 数の平均値。
正規化 RT 数	Tweet に対する共感の有無とは無関係に, 他人からの注目度合いを示す指標。Tweet の RT 数をそのまま使用すると, 発言したアカウントの普段の RT 数を考慮していないため, 当該 Tweet がこのアカウントから発信されたものとして RT が多かったのか少なかったのかを示す値とならず, 属性として不適切である。これを解消するため, 取得した RT 数を, アカウントの平均 RT 数で割ることで正規化した値。普段の RT 数の X 倍を表す。
正規化 Fav 数	他人からの共感の度合いを示す指標。取得した Fav 数を, アカウントの平均 Fav 数で割ることで正規化した値。普段の Fav 数の X 倍を表す。正規化の必要性は上記の正規化 RT 数と同じである。
正規化 Tweet 数	アカウントがどれ程活発かを示す指標。取得した Tweet 数を, アカウントの Twitter 歴 (日換算) で割ることで正規化した値。1 日あたりの Tweet 数を表す。正規化の必要性はアカウントごとの Twitter 歴の長短の影響を吸収するためである。
発言の極性	炎上元の Tweet(発言) に対し肯定 (P)・否定 (N)・中立 (E) の極性タグを付与。
(平均) 味方率	ある Tweet の ReTweeter のうち, 発言者の Follower が占める割合。ReTweeter の中にどれだけ味方がいるかを示す。平均味方率は, アカウントごとの味方率の平均値。味方率が高いと, Follower が味方となって庇ってくれるので炎上しにくく, 逆に低いと, 味方がいないので炎上しやすいという仮説に基づいている。
有名人/非有名人	Wikipedia に接続し, その名前の記事があれば有名人, なければ非有名人とする。
炎上/非炎上	該当 Tweet が炎上元となって炎上しているかどうかを示す。1.1 の炎上の定義と, インターネット上で炎上記事になっている [フィーノ社 02, ソシオコーポレーション 06] ことを元に, 人手でタグ付けする。決定木ではこの属性を目的変数として分類を行う。

4.2 炎上収集モデルの学習

本稿では、RT 数の多い Tweet に炎上 Tweet が含まれている可能性があると考え、以下のデータ収集プロセスに従って、炎上収集モデルを学習するためのデータを収集した。

- (1) NAVER 炎上まとめ [NAVER 10] に記載の SBCV41 件のうち、炎上を引き起こした Tweet が削除されていない 20 件を手手で収集する。
- (2) (1) を発生させたアカウントを対象アカウントとする。
- (3) 対象アカウントそれぞれの最新 Tweet 3200 件を抽出する。この値は TwitterAPI の上限である。
- (4) Tweet 日 2013/01/01 ~ 2013/06/01 の半年間以外を除去することで、Tweet の収集期間を整える。
- (5) 各アカウントに対し、抽出した各 Tweet の RT 数を平均し、アカウントの平均 RT 数を計算する。
- (6) 各アカウントの普段とはかけ離れた盛り上がりを表現するため、計算した平均 RT 数を経験的に 40 倍し、この閾値を超えた Tweet を収集する。その結果、炎上 17 件、非炎上 106 件の計 123 件が得られた。
- (7) (1) の 20 件と (6) の 123 件を合わせた 143 件を学習用データとする。

こうして得られた学習用データに対し、前節で挙げた全属性を使い、Weka [Hall 09] に含まれる J48 で収集モデルを作成した。

作られた決定木を図 5 に示す。収集モデルの観察により次のような考察が得られる。

- 有名人/非有名人が木に現れず、一方 Follower 数の多さが上位の分類属性となっていることから、Wikipedia に掲載されているかどうかで判定できる有名人/非有名人という概念は炎上判定には有効ではなく、むしろ Twitter 独特のネット有名人/ネット非有名人という区分が有効であると考えられる。
- 図の左側、Follower 数と味方率が少ないと炎上するという部分はネット非有名人の炎上を示していると考えられる。
- 図の右側、Follower 数が多く、Fav 数と Follow 数が少なく、かつ平均味方率が約 1 割以下だと炎上するという部分は、ネット有名人の炎上を示している読み取れる。
- この考察から、炎上収集においては味方率が有効に機能することが示された。また (平均) 味方率の少なさは対象アカウントのファンの少なさを示し、炎上の要因の一つであると考えられ、これについては 6 章で議論している。

4.3 炎上の収集

炎上収集モデルを使ってより多くの炎上 Tweet を得る手順を説明する。

- (1) Twitter の日本における Follower 数ランキングの

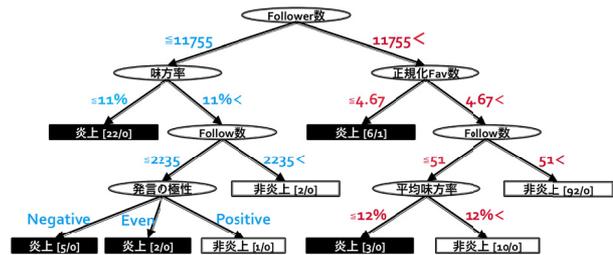


図 5 炎上収集モデル

上位 5000 を対象アカウントとして、4.2 節のデータ収集プロセスの 3 ~ 6 を介して Tweet 1533 件を収集する。

- (2) 抽出した Tweet に対し、Bot による投稿・Blog や公式サイトの更新情報・画像や動画付き Tweet・日本語以外の Tweet を目視で除去した計 235 Tweet を炎上 Tweet 収集対象データとする。
- (3) 炎上 Tweet 収集対象データに対し、収集モデルで分類を行ったところ、炎上 Tweet 24 件、非炎上 Tweet 211 件となった。真の炎上 Tweet に関して調査したところ、適合率 100%、再現率 93%であった。

5. 実験

5.1 未来の炎上を予測する決定木モデル

決定木を用いた炎上収集は成功したが、味方率という投稿後でなければ計算できない属性を含んだモデルであり、炎上予測を行うためには、投稿前に得られる属性だけで分類を行う必要がある。そこで、味方率に代えて Tweet 投稿時点での dcndp を新たに加える。炎上予測モデルで用いる属性は、発言の極性・平均 RT 数・平均 Fav 数・平均味方率・正規化 Tweet 数・Follow 数・Follower 数・有名人/非有名人・割引累積日次極性 (dcndp) の 9 種である。

5.2 予測結果と考察

4.2 節で説明した収集モデルの学習データ 143 件と 4.3 節で説明した炎上収集対象データ 235 件を合わせた 378 件を用い、炎上予測モデルの学習と評価を行う。ただし、本実験では以下の条件を満たしたデータのみを用いる。

- Tweet のトピックが人物など単一の固有表現と明確に断定できること。これは 3.4 節で示した方法により dcndp を計算するためである。
- SBCV 発生メカニズム仮説と無関係に議論を呼びがちなトピックを持たないこと。本実験では、ジェンダー・政治思想・宗教・社会/国際問題をトピックとして持たないものとした。

さらに、データの炎上/非炎上のバランスを取るため、非炎上データを機械的に 1/3 に間引いた。その結果、炎上 Tweet 22 件、非炎上 Tweet 32 件の計 54 Tweet が得ら

表 3 炎上予測モデルの精度

	Precision	Recall	F-Measure
炎上	0.952	0.909	0.930
非炎上	0.939	0.969	0.954
加重平均	0.945	0.944	0.944

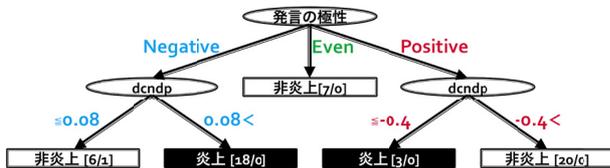


図 6 炎上予測モデル

れた。これを J48 に与え、10 重交差検定で精度を測ったところ、94%の精度で炎上を予測できた(表 3)。

作られた決定木(図 6)から次のような考察が得られた。否定的な発言のうち世評を表す dcndp が +0.08 以上、または肯定的な発言のうち世評を表す dcndp が -0.4 以下だと炎上する。これにより発言の評価と世評の対立が炎上を導くという作業仮説の正しさがデータにより裏付けられたとともに、炎上データが否定的な発言側に偏っていることから、肯定的な発言より、否定的な発言の方が炎上しやすいことがわかった。

なお、2.2 で述べたように、分析結果の経時変化に興味を持たない評判分析ではトピックに関する過去の累積的態度に着目することになるため、そうした手法により数値化された世評を dcndp の代わりに用いることは可能と考えられ、その場合を本章の結果と比較することは今後の課題である。累積的態度を表す素性としての dcndp の特徴は忘却係数を用いていることである。

6. 味方率の応用

4 章で説明した味方率の応用として、味方率マネジメントを提案する。味方率の高い/低いアカウントを特徴付けるため、Follower 数上位 500 アカウントにおいて、平均味方率 70%以上、30%以下の各アカウントに対し人手で分類タグ*2を付け、その頻出タグをワードクラウド化 [Chien 11] した。平均味方率が 70%以上のアカウントは 500 アカウント中 105 アカウント(図 7)、平均味方率が 30%以下のアカウントは 500 アカウント中 74 アカウントだった(図 8)。

平均味方率が高いアカウントの分類としては、AKB・アーティスト・ニコニコ歌手・声優・女優などが多く見られた。これらはファンが多いため、非難を受ける可能性が低く、炎上しにくいことが予想される。また、こ

*2 アーティスト、一般人、詩人、漫画家、声優、女優、実業家、芸人、評論家、広報、bot など。



図 7 平均味方率 70%以上のアカウントまとめ



図 8 平均味方率 30%以下のアカウントまとめ

れらを批判することで多くのファンを敵に回し、逆に炎上させられる可能性がある。

平均味方率が低いアカウントの分類としては、芸人・評論家・起業家・政治家・漫画家・Artist などが多く見られた。これらはファンが少ないため、非難を受けやすく、炎上しやすいことが予想される。

ここで、平均味方率の低さを目安に、アカウントのあり方を再考させるマネジメントを提案する。味方率が得られないとは、Tweet が自分の Follower に応援されていないということなので、よりファンサービスして味方を増やすべき、というようにマネジメントに応用できる。

今後の課題として、個別の人物について味方率の高い、低いがあった時に味方率を左右する要素は何なのか、を調査することが考えられる。Wikipedia の記述を味方率の高い人、低い人に分類することで味方の多い(少ない)人の特徴を読み取れる可能性がある。

7. 関連研究

7.1 炎上に関連する研究

炎上に関連する研究として以下のようなものがある。山本ら [山本 09] は CGM における 150 件のトラブル事例からキーワードを抽出、タグ付けを行うことで炎上事例の調査を行った。田代 [田代 11] はインターネットトラブルを「金銭トラブル」「コミュニケーショントラブル」「情報管理トラブル」「心身トラブル」という 4 種類に分類した。本研究の炎上は、この分類中の「コミュニケーショントラブル」に相当する。また(株)プラスアルファコンサルティング [PlusAlphaCo. 11] では、投稿前に責

任者へ投稿可否問い合わせメールを自動送信することで、炎上しそうな投稿を防ぐシステムが開発されている。

7.2 Twitter による社会事象の予測に関連する研究

Twitter 分析による社会事象の予測としての類似研究との差異を述べる。Bollen らは、Twitter から取得した世評から株価を予測する方法を提案した [Bollen 10]。当該論文では株式市場という社会性の高い対象の動きを予測するのに対し、本研究では個人の Tweet というパーソナルな対象の社会への受け取られ方を予測するという点で異なる。また、Bollen らは世評が株価に影響するメカニズムには触れない立場であるのに対し、本研究では SBCV 仮説を提示し、それをデータで裏付ける構成的立場を取っているため、Tweet しようとする内容の炎上予測結果に基づき炎上しないための具体的なアクションを取ることができる。Bollen らは米国大統領選などの社会的大事件が Twitter 上の世評にどう影響するかも調べている [Bollen 11] が、社会事象を用いた世評の予測であり、予測に用いる材料と予測される事象の関係が異なる。

7.3 世評の時間的变化に関連する研究

本論文で提案する動的評判分析と既存の時間に伴う極性変化に関する研究の違いについて述べる。Nguyen らは、あるトピックに関する世評がどう変化するかをサポートベクターマシンで予測する方法を提案した [Nguyen 12]。上記研究で用いられている世評の指標である Sentiment Ratio は ndp と類似の式で定義されるが、 ndp で考慮している中立 Tweet 数を Sentiment Ratio は考慮していないこと、上記研究では Sentiment Ratio のみで世評を可視化しようとしているが本研究では可視化において ndp だけでなく肯定・否定 Tweet 絶対数も表示しておりわかりやすいことが相違点として挙げられる。さらに、RT 数など Tweet のメタ情報や、Follower 数などアカウントから得られる素性を、世評ではなく個々の Tweet の評価に使用している点も本研究の特徴である。また、Bifet らは、Tweet を tf-idf に基づきモデル化し、顔文字を肯定/否定の教師データに用いて決定木学習することで Tweet の極性を判定、その結果に基づき Tweet の肯定率の時間的推移を可視化した [Bifet 11]。当該論文に対し、本研究では肯定/否定判定に学習が不要な点、極性のうち中立を考慮している点が特徴である。

7.4 頻出語出現量の差を用いた手法 [石野 12]

2.4 で触れた特徴語抽出のアルゴリズムである頻出語出現量の差を用いた手法について説明する。本手法は、あるコーパスの頻出語からそれを含むより大きなコーパスの頻出語を差し引くことにより当該コーパスの特徴語を抽出する手法である。2.4 (4) では、トピック「チノパン」に関する、ある特定期間の Tweet の頻出語からそれを含む収集データ全期間の Tweet の頻出語を差し引くことで

特定期間の特徴語を抽出している。

7.5 P/N 判定 [Turney 02]

P/N 判定とはトピックを肯定と否定の 2 極性に分類することである。Turney は、コーパスから得られる共起情報を利用し、肯定的な極性を持つ代表的な表現「good」や「excellent」と、否定的な極性を持つ代表的な表現「bad」や「poor」を種表現として予め用意しておき、種表現と共起する比率にしたがって語句の評価極性値を判定した。この手法は少ない言語資源から多量の P/N 判定情報を獲得することができる。

本研究は、P/N 判定を発展させた、動的な P/N 判定に関するものと位置づけられる。

7.6 ハイブリッド分類器 [Mullen 04]

Mullen らは、SVM 分類器を用いたセンチメント分析システムに、Turney の極性値や WordNet から抽出される極性値を素性として追加したハイブリッド分析システムを提案した。実験の結果、素性の追加により精度に改善が見られた。

8. ま と め

本稿において、CGM における炎上を定義し、炎上を同定するための手法を提案した。まず、炎上事例から、犯罪自慢・SBCV・暴露という 3 つの炎上様式を定義した。この様式のうち、その予測にニーズがあると思われる SBCV を主題に議論を進めた。世間の評価と対立する評価を含む発言で炎上するという作業仮説に基づき事例を分析した上で、過去の Twitter データに対し、トピックに対する世間の評価やアカウントの特性を数量化した属性による決定木モデルを用いた炎上予測手法を提案した。その結果、Twitter 上での SBCV 型の炎上予測できることを示した。結果として得られた決定木モデルには Twitter 独自の属性は含まれておらず、発言の極性と、何に対して言及しているかがわかれば炎上予測が可能のため、ブログ記事や TV での発言など、本稿で取り扱った Twitter 以外のメディアにも適用できる可能性がある。

謝 辞

本研究は、JSPS 科研費 24300005, 23500039, 25730038 の助成を受けたものです。本研究を遂行するにあたり、研究の機会と議論の場の提供、活発な議論と貴重な御意見を頂いた国立情報学研究所/東京大学の本位田研究室の皆様、早稲田大学の深澤研究室の皆様へ感謝致します。

◇ 参 考 文 献 ◇

- [Bifet 11] Bifet, A., Holmes, G., Pfahringer, B., and Gavalda, R.: Detecting Sentiment Change in Twitter Streaming Data, *Journal of Machine Learning Research*, pp. 5–11 (2011)

- [Bollen 10] Bollen, J., Mao, H., and Zeng, X.-J.: Twitter mood predicts the stock market, *Journal of Computational Science*, Vol. 2, No. 1, pp. 1–8 (2010)
- [Bollen 11] Bollen, J., Mao, H., and Pepe, A.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena, in *ICWSM'11*, pp. 450–453 (2011)
- [Chien 11] Chien, T.: HTML5 Word Cloud (2011), <http://timc.idv.tw/wordcloud/>
- [Hall 09] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H.: The WEKA Mining Software: An Update, in *SIGKDD Explorations*, Vol. 11, pp. 10–18 (2009)
- [Kudo 04] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, in *EMNLP-2004*, pp. 230–237 (2004)
- [LivedoorNEWS 13] LivedoorNEWS, : パイオリニスト AKB を侮辱し炎上 (2013), <http://news.livedoor.com/topics/detail/8032049/>
- [Mullen 04] Mullen, T. and Collier, N.: Sentiment analysis using support vector machines with diverse information sources, in *EMNLP-2004*, pp. 412–418 (2004)
- [NAVER 10] NAVER まとめ: 犯罪自慢・炎上まとめ (2010), <http://matome.naver.jp/odai/2132708118341913001/>
- [Nguyen 12] Nguyen, L., Wu, P., Chan, W., Peng, W., and Zhang, Y.: Predicting collective sentiment dynamics from time-series social media, in *WISDOM-2012* (2012)
- [PlusAlphaCo. 11] PlusAlphaCo., : カスタマーリングス (2011), http://www.pa-consul.co.jp/LP_rings_mail/
- [Togetter 13] Togetter, : 千本桜は AKB の曲です (2013), <http://togetter.com/li/551523>
- [Turney 02] Turney, P. D.: Thumbs up? thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews, *ACL-2002*, pp. 417–424 (2002)
- [Twitter 13] Twitter, : Using the Twitter Search API (2013), <https://dev.twitter.com/docs/using-search/>
- [ソシオコーポレーション 06] ソシオコーポレーション : ロケットニュース 24 (2006), <http://rocketnews24.com/>
- [フィーノ社 02] フィーノ社: 探偵ファイル (2002), <http://www.tanteifile.com/>
- [井上 07] 井上 結衣, 藤井 敦: 時事問題に関する賛否両意見の収集, 情報処理学会研究報告. 自然言語処理研究会報告, pp. 93–98 (2007)
- [奥村 05] 奥村 学, 難波 英嗣: 知の科学 テキスト自動要約, オーム社 (2005)
- [乾 06] 乾 健太郎, 奥村 学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol. 13, No. 3 (2006)
- [岩崎 13] 岩崎 祐貴, 折原 良平, 清 雄一, 中川 博之, 田原 康之, 大須賀 昭彦: CGM における炎上の同定とその応用, 合同エンジニアワークショップ & シンポジウム 2013 (JAWS 2013) (2013)
- [鯨岡 11] 鯨岡 孝子: 性的虐待を隠匿した疑いのフットボールコーチをかばう発言で炎上 (2011), <http://news.livedoor.com/article/detail/6018556/>
- [山口 13] 山口 香: 谷亮子の五輪代表選出に憤慨し提訴考えたこともあった (2013), 女性セブン, 2013 年 2 月 28 日号.
- [山本 09] 山本 太郎, 畑島 隆, 谷本 茂明, 高橋 克巳: CGM におけるトラブル事例の整理と調査手法について, 情報処理学会研究報告, 第 2009 巻, pp. 103–108 (2009)
- [小林 05] 小林 のぞみ, 乾 健太郎, 松本 裕治, 立石 健二, 福島 俊一: 意見抽出のため評価表現の収集, 自然言語処理, Vol. 12, No. 2, pp. 203–222 (2005)
- [小林 11] 小林 直樹: ソーシャルメディア炎上事件簿, 日経デジタルマーケティング (2011)
- [石野 12] 石野 克徳, 折原 良平, 中川 博之, 田原 康之, 大須賀 昭彦: フォークソノミーとソーシャルアノテーションを用いた動画共有サービス利用支援の試み, 情報処理学会論文誌, Vol. 53, No. 11, pp. 2494–2506 (2012)
- [田代 08] 田代 光輝: ブログ炎上 学びとコンピュータハンドブック, 東京電機大学出版局 (2008)
- [田代 11] 田代 光輝: インターネットトラブルの分類方法の提案, 情報社会学会誌, 第 6 巻, pp. 101–114 (2011)
- [東スポ 13] 東スポ : 軽率ツイートで楽しんで 芸人生命の危機 (2013), <http://www.tokyo-sports.co.jp/entame/>

69832/

[東山 08] 東山 昌彦, 乾 健太郎, 松本 裕治: 述語の選択好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, pp. 584–587 (2008)

[担当委員: 井口 誠]

2014 年 5 月 8 日 受理

著者紹介

岩崎 祐貴



2014 年電気通信大学大学院情報システム学研究科修士課程修了。同年 (株) サイバーエージェント入社。主として機械学習・自然言語処理を用いたマイクロブログの分析に從事。

折原 良平 (正会員)



1988 年筑波大学大学院工学研究科電子・情報工学専攻修士前期課程修了。同年, (株) 三菱総合研究所入社。現任, 同社研究開発センター知識メディアラボラトリーに勤務。1993-95 年 University of Toronto, Department of Industrial Engineering 客員研究員。2010 年より電気通信大学情報システム学研究科客員教授。発想支援技術, 類推, 機械学習, データ・テキストマイニングの研究に從事。2009 年度人工知能学会論文賞, 2010 年度人工知能学会功労賞, 2012 年度情報処理学会活動員献賞受賞。博士 (工学)。

清 雄一



1981 年生。2009 年東京大学大学院情報理工学系研究科修士後期課程修了。同年 (株) 三菱総合研究所入社。同社情報技術研究センター, 金融ソリューション本部等に所属。2013 年より電気通信大学大学院情報システム学研究科助教, 現在に至る。分散コンピューティング, セキュリティ, プライバシー保護技術等の研究に從事。情報処理学会, 電子情報通信学会, IEEE Computer Society 各会員。

中川 博之



1974 年生。1997 年大阪大学基礎工学部情報工学科卒業。同年鹿島建設 (株) に入社。2007 年東京大学大学院情報理工学系研究科修士課程修了, 2008 年同大学院博士課程中退。同年電気通信大学助教, 2014 年大阪大学大学院情報科学研究科准教授, 現在に至る。工学博士 (早稲田大学)。要求工学, 形式手法, エージェントおよび自己適応システム開発手法の研究に從事。情報処理学会, 電子情報通信学会, IEEE CS 各会員。

田原 康之



1966 年生。1991 年東京大学大学院理学系研究科数学専攻修士課程修了。同年 (株) 東芝入社。1993–96 年情報処理振興事業協会に。1996–97 年英国 City 大学客員研究員。1997–98 年英国 Imperial College 客員研究員。2003 年国立情報学研究所着任。2008 年より電気通信大学准教授。博士 (情報科学)(早稲田大学)。エージェント技術, およびソフトウェア工学などの研究に從事。情報処理学会, 日本ソフトウェア科学会各会員。

大須賀 昭彦 (正会員)



1958 年生。1981 年上智大学理工学部数学科卒。同年 (株) 東芝入社。同社 研究開発センター, ソフトウェア技術センター等に所属。1985–89 年 (財) 新世代コンピュータ技術開発機構 (ICOT) 出向。2007 年より, 電気通信大学大学院情報システム学研究科教授。2012 年より, 国立情報学研究所客員教授兼任。工学博士 (早稲田大学)。主としてソフトウェアのためのフォーマルメソッド, エージェント技術の研究に從事。1986 年度情報処理学会論文賞受賞。IEEE Computer Society Japan Chapter Chair, 人工知能学会理事, 日本ソフトウェア科学会理事を歴任。情報処理学会, 電子情報通信学会, 人工知能学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員。