

CGMにおける炎上の同定とその応用

Identification of Flaming and Its Applications in CGM

岩崎祐貴^{*1} 折原良平^{*1} 清雄一^{*1} 中川博之^{*1} 田原康之^{*1} 大須賀昭彦^{*1}
 Yuki Iwazaki Ryohei Orihara Yuichi Sei Hiroyuki Nakagawa Yasuyuki Tahara Akihiko Ohsuga

^{*1} 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

In recent years, thanks to the spread of Consumer Generated Media, anybody can easily express their opinion publicly. Because of this, a phenomenon of flooding criticism on the Internet, called *flaming*, frequently occurs. There is an increasing risk of suffering damage for not only a celebrity, but also ordinary people. This research aims to prevent a flaming. Concretely, we propose methods to identify flamings on Twitter. First, we tried to extract typical words used by *flaming recidivists* to identify their flamings. However, this approach yielded a poor result. Second, we divided flamings into two categories, namely, *criminal episode* and *different values*. For the criminal episodes, an identification method based on Naïve Bayes was successfully performing the task. We are tackling an issue to identify the different values. A method to focus on the deviation of one's opinion from the consensus of the community is promising.

1. 背景と目的

近年、ブログや twitter といった CGM(Consumer Generated Media)の普及により誰でも気軽に情報発信できるようになった。それに伴いインターネット上で批判が殺到する炎上という現象が頻繁に起こり、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。

炎上をテーマとした既存研究は存在するが、機械的に炎上予測を行い、人間の介入なく炎上防止を行うシステムの例はない。

そこで本研究では、炎上事例を抽出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的とする(図 1)。

2. 炎上とは

炎上とは、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状況と定義する[田代 08]。単純にコメントが多い状況、特に批判的なコメントだけでなく肯定的なコメントも入り混じった賛否両論の議論が盛り上がっている状況は、本稿では炎上とは呼ばない。

次節から、この炎上をどう抽出するかを、仮説を立てて論じていく。

3. 仮説 A

本研究での目的である炎上を同定するために炎上常習犯の発言から炎上特徴語が抽出できる、ということを仮説として挙げる。

3.1 仮説 A 検証実験

常習犯の発言内容を限定した方が特徴の抽出をしやすいと考え、炎上事例 3 件以上の政治家 3 名の最新 tweet3200 件に対し、webidf を重みとして特徴語を抽出する。webidf とは単語がどれくらい特徴的かの指標であり、今回は頻出語出現量の差を用いる石野らの手法[石野 11]を適用した。

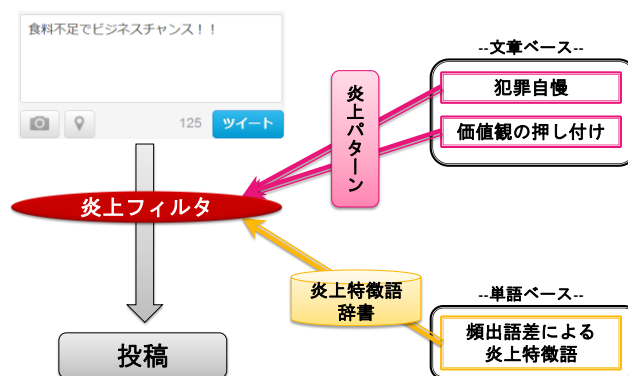


図 1: システム概要図

3.2 実験結果と考察

本手法の適用前後で webidf が平均で 34% 増加したことから、一般的な単語をある程度排除できたことがわかった。特徴語としては「アホ・バカ・無理・意味不明」など批判的な語が抽出された。しかし、こういう乱暴な語を使わない発言でも炎上することがあり、炎上特徴語としては弱いと考えられる。

結果として、炎上事例をひとまとめにして非炎上事例と比較して特徴語を抽出し、それに基づいて炎上事例を同定する方法ではうまくいかないことがわかった。

4. 仮説 B

3.2 で述べた通り、仮説 A では不十分であるということがわかった。参考文献[小林 11]によると炎上は数種類に類型化できるとされており、実際に 100 件の炎上事例を人手で分類したところ、犯罪自慢と価値観の押し付けによる炎上がほとんどを占める結果となった(表 1)。このことから、炎上は 2 種類でパターン化できるということを新たに仮説として挙げる。そしてこの 2 種類の炎上の抽出方法を提案する。

表 1. 炎上事例分類結果(N=100)

犯罪自慢	価値観の押し付け	その他
51.0%	41.0%	8.0%

4.1 犯罪自慢

犯罪自慢によって起こる炎上のことで、一般人の炎上が圧倒的に多い。「無免許」「飲酒運転」「〇〇に爆弾を仕掛けた」等、犯罪自慢特有の単語が目立つという特徴がある。そのことから、spamメールの抽出方法と同じベイジアンフィルタを用い、犯罪自慢かそうでないかを判定した。

犯罪自慢 tweet100 件、一般 tweet300 件を実験データとして、これをさらに学習用、テスト用に 4:1 で分け、5 重交差検定を行った。その結果、高い精度で犯罪自慢を抽出できることがわかった(表 2)。

表 2. 抽出精度評価結果

再現率	適合率	F 値
96.5%	70.2%	81.2%

4.2 価値観の押し付け

世間での評価(世評)とずれた評価を含む発言によって起こる炎上のことで、著名人の炎上が多い。「紀子様出産は別にめでたくない」が例である。天皇家の紀子様のご出産をお祝いする風潮が一般的である中、めでたくないという否定的な発言、つまり世評とずれた評価をしてしまい炎上したと思われる。

別の例として、「園田監督は素晴らしい人だ」がある。これは、体罰問題で園田監督に批判が集中している中、素晴らしいという擁護する発言をしてしまい炎上した例である。

5. 関連研究

5.1 炎上に関連する研究

炎上に関連した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、株式会社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。

5.2 頻出語出現量の差を用いた手法[石野 11]

仮説 A の実験で用いた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。本実験では、炎上の多い政治家、炎上のない一般の政治家ともに政治に関連した単語が特徴語として現れることを利用し、炎上常習犯頻出語から一般政治家頻出語を差し引くことで政治家に共通する特徴語を排除し、炎上の多い政治家独自の特徴語が抽出されることを狙っている。

5.3 P/N 判定[橋本 10]

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。

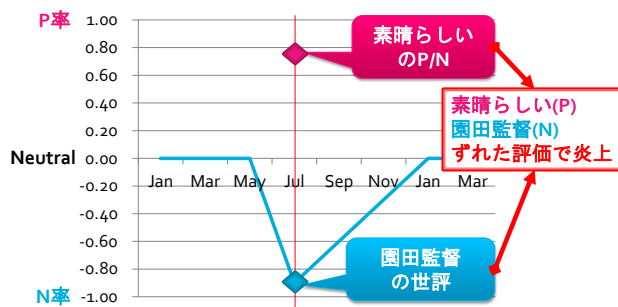


図 2: ずれた評価による炎上

4.2 で述べた事例について考える。園田監督というトピックの極性は、体罰問題以前には中立的であったが、体罰問題以降は否定に転じたと考えられる。すなわち、トピックの極性は時間的に変化するもの[東 11]であり、その変化にはニュースが大きく影響している。

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(図 2)。

6. まとめと今後

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仮説を立てた。そしてその仮説に対し検証分析を行い、結果として、仮説 A の炎上常習犯に注目した炎上特徴語の抽出はうまくいかず、新たに炎上は 2 種類でパターン化できるという仮説 B を立てた。

現在までに、炎上パターン例の 1 種である犯罪自慢に関してベイジアンフィルタを用い、高い精度で抽出できたという成果が出ている。今後はパターン例のもう 1 種である価値観の押し付けに対し、炎上する語やフレーズの組み合わせの同定方法を調査していく。また、炎上単語として、差別発言が多く含まれている放送禁止用語辞書や、橋本ら[橋本 10]の作成した P/N 辞書の使用も考慮している。さらに、5.2 で挙げた動的な P/N トピックを抽出するシステムの実装を検討していく。

参考文献

- [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
- [石野 11] 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シーンへのラベリング」, 第 10 回合同エージェントワークショップ & シンポジウム 2011(JAWS2011)
- [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
- [山本 09] 山本 太郎, 畑島 隆, 谷本 茂明, 高橋 克己:「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, 2009
- [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, 2011
- [PlusAlphaCo 11] 「カスタマーリングス」, http://www.pa-consul.co.jp/LP_rings_mail/
- [橋本 10] 橋本 和幸, 中川 博之, 田原 康之, 大須賀 昭彦:「センチメント分析とトピック抽出によるマイクロブログからの評判傾向抽出」, 第 9 回合同エージェントワークショップ & シンポジウム 2010(JAWS2010)
- [東 11] 東 浩紀:『一般意志 2.0 ルソー、フロイト、グーグル』, 講談社, 2011