

機械学習を利用した SNS における炎上画像の推定

橋本 山河† 永井 由佳里†

北陸先端科学技術大学院大学 知識科学研究科‡

1. はじめに

近年, 炎上という言葉が注目を集めている[1]. 炎上とは, Twitter や Facebook などのソーシャルネットワークサービス(SNS)へ投稿した内容に批判的なコメントが殺到する状況である[2].

炎上が社会に与える影響は大きい. 従業員の Twitter への投稿が炎上したことにより, 店舗が閉店する事態になった事例がある. SNS への不適切な写真の投稿が報道されたことで, 炎上が目目されるようになったと考えられている[1].

炎上を未然に防ぐ方法として, SNS へ投稿される前に内容を機械的に検閲し, 炎上の可能性を有する投稿を防ぐシステムが提案されている[3]. 提案されているシステムは投稿文を分析し, 炎上フィルタリングをおこなうものである.

本研究の目的は, 画像分析による炎上検知の検討と, 炎上フィルタリングへの応用である. 本研究では2つの手法の機械学習を試み, 炎上の分類性能を検討した.

2. 関連研究

炎上の同定に関する研究として次のようなものがある. 岩崎ら[3]は, 事前に炎上した投稿を犯罪自慢, 価値観の押し付け, その他に分類し, 犯罪自慢に含まれる投稿文をベイジアンフィルタによって抽出している.

Yoshida ら[4]は, 企業への SNS のコメントに対し, ポジティブ, ニュートラル, ネガティブのラベルを付与した後, 機械学習を行い, ネガティブに分類されるコメントの量によって炎上の同定を行っている.

炎上の同定に関する研究の多くはテキストを用いた機械学習であり, 画像による炎上分析に注目したものは少ない. Twitter へ「○○(店名) なる」という内容の投稿文と, 飲食店内で自身の

鼻に醤油差しを刺した状態を撮影した画像を投稿した結果, 炎上した事例があり, 既存のテキストによるフィルタリングでは不十分である.

3. 炎上画像の分類と分析

本研究では, 炎上した投稿と同時に投稿された画像を炎上画像とし, それ以外の画像を非炎上画像とする.

炎上画像と非炎上画像の分類を行うにあたり, 岩崎ら[3]と同様に, 事前に炎上画像の種類の分類を行った. 分類結果は, 店舗の醤油差しを鼻に挿すような”挿す炎上”, コンビニエンスストアの冷蔵庫や買い物カゴなどに入るような”入る炎上”, 線路の上や食品の上に寝そべる写真などが含まれる”上に乗る炎上”, 上記3種類に含まれない”その他炎上”となった. 分類件数と割合を表1に示す.

入る炎上	挿す炎上	上に乗る炎上	その他炎上	合計
11	21	18	7	57件
19.3%	36.8%	31.6%	12.3%	100.0%

表1 炎上画像の分類と割合

炎上画像の分析を以下の2つの手法で行った.

手法1は, はじめに画像中の物体を人手によって抽出を行う. 抽出された物体を特徴量とし, サポートベクターマシンによって分類する.

手法2は, はじめに画像中から SURF 特徴量を抽出する. 抽出処理によって, 1枚の画像から128次元の特徴量が複数抽出される. 次に k-means を適用し, 各炎上の種類ごとに代表される特徴量を求める.

1枚の画像中に含まれる代表的な特徴量の数を計測した値を手法2における最終的な特徴量とする. 最終的な特徴量を手法1と同様にサポートベクターマシンで分類する.

手法2において, 学習に使用したデータを用いて炎上画像の予測を行った結果, 炎上画像と非炎上画像を正しく分類できていたが, 炎上の種類に

Estimation of images in flaming SNS using a machine learning algorithm.

Hashimoto Sanga† Nagai Yukari†

‡Japan Advanced Institute of Science and Technology

関しての分類性能をもっていなかった。そのため、手法2の分類を、炎上か非炎上かの2種類の分類にした。

4. 手法の評価

各手法の分類手法の結果を表2, 表3に示す。

入る炎上	挿す炎上	上に乗る炎上	その他炎上	炎上検知率	分類精度
80.0%	100.0%	80.0%	0.0%	72.2%	92.6%

表2 手法1の性能

炎上画像検知率	非炎上画像反応率	分類精度
66.7%	40.0%	61.7%

表3 手法2の性能

表2の炎上検知率は、入る炎上・挿す炎上・上に乗る炎上・その他炎上・非炎上画像の分類を正しく予測した割合である。分類精度は、炎上画像か非炎上画像かを正しく予測した割合である。

分類精度は92.6%で分類されている。炎上の種類を含めた精度の、炎上検知率は72.2%であった。

各種類の炎上画像に多く含まれていた上位の物体が表4である。全ての炎上画像に人が含まれていた。冷蔵庫や食品といった、食事に関連する物体と、線路やホームなど鉄道に関連する物体が炎上画像に含まれていることが多い。

入る炎上		挿す炎上	
物体	出現率	物体	出現率
人	100.0%	人	100.0%
冷蔵庫	81.8%	食器	71.4%
食器	63.6%	食品	61.9%
上に乗る炎上		その他炎上	
物体	出現率	物体	出現率
人	100.0%	人	100.0%
線路	77.7%	台所	42.8%
ホーム	50.0%	食品	42.8%

表4 炎上画像中に移っている物体

手法2の分類精度は61.7%と高くない。k-meansによって各種炎上、非炎上の特徴量の代表される特徴量抽出時に、似通った特徴量を抽出している可能性がある。

炎上画像検知率は炎上画像に対し炎上の予測ができた割合である。非炎上画像反応率は、非炎上画像に対し炎上の予測を誤って予測した割合である。

手法2は、炎上画像検知率66.7%と非炎上画像反応率40.0%と精度が高くないことから、炎上画像

の推定に最適な手法であるとは言えない。

炎上画像予測のための特徴量として SURF 特徴量を1つの指標として取り込むことで予測精度の向上に繋がると考えられる。

5. まとめと今後

近年、注目されている不適切画像による炎上に対し、機械学習を用いたフィルタリングに応用するために、画像分析による、炎上画像と非炎上画像の分類を試みた。

分析の結果、人手によって抽出した物体を特徴量とした場合、92.6%で分類できた。現状では物体検出を人が行うため、機械的に炎上をフィルタリングすることはできない。しかし、人手による物体抽出部分は、一般画像認識の精度が高くなることで、人手を利用しない物体検出が可能になり、炎上画像の推定が可能になると思われる。

SURF 特徴量を用いた場合は、精度が高くないものの、炎上画像に反応しやすく、非炎上画像への反応を低下させている。画像中の物体や SURF 特徴量を使用した分析によって、炎上検知の精度向上に貢献していることから、炎上画像にはなんらかの特徴がある可能性が示されている。

手法2には SURF 特徴量を抽出した後の k-means によって、炎上特有の特徴量が抽出されない可能性があるため、非炎上画像に含まれている特徴量は削除する処理など、最適化の余地が残されている。

本研究における分析はいずれも過去の炎上画像を利用しているため、未知の炎上を分類することができない。そのため、未知の炎上に対応する炎上画像分析などを加え、予測精度の向上にむけて今後も継続して炎上画像の分析を行う予定である。

参考文献

[1] 総務省, “平成 26 年版情報通信白書”, 総務省, 2014.
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc143120.html> 2016/01/04 Access

[2] 田代光輝, “大学生のネット炎上分析と予防及び対応の提案 : 好ターゲットとしての大学生の実情とネット炎上からの回避の提案”, 大妻女子大学紀要, 2012

[3] 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”, 人工知能学会論文誌, 2013

[4] Shun Yoshida, Jun Kitazono, Seiichi Ozawa, Takahiro Sugawara, Tatsuya Haga, Shogo Nakamura, “Sentiment analysis for various SNS media using Naive Bayes classifier and its application to flaming detection”, In Proc. of 2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD), 2014