第29回年次大会予稿

炎上の早期解決を促す情報解析システムの実現のため,発言の立場ごとに発言内容を分類する手法の研究

Research on a method to classify the content of statements according to the position of the statement in order to realize an information analysis system that promotes early resolution of flaming

横山拓也¹, 箕輪弘嗣^{1*}, 渡邊憲二¹

Takuya YOKOYAMA¹, Hirotsugu MINOWA^{1*}, Kenji WATANABE¹

¹ 岡山商科大学

Okayama Shoka University

〒700-0087 岡山県岡山市北区津島京町2丁目10-1

E-Mail: minowa@po.osu.ac.jp

*連絡先著者 Corresponding Author

近年,SNSは若い人を中心に普及し,それに付随して,企業も率先して利用しているメディアである.しかし,SNSでは炎上と呼ばれる激しい議論が生じてしまう事例がある.本研究では,炎上の早期解決を支援するためのシステムの開発を目指している.本論文では,システム実現のため,Twitterで炎上したツイートのスレッドを対象に複数の機械学習法を用いて意見の賛成派,反対派の分類を機械学習で試みた結果,1例だけだが,86.6-88.1%の精度で分類できた事について報告する.

In recent years, SNS has become widespread mainly among young people, and along with that, companies are also taking the initiative in using it. However, there are cases in which a fierce debate called flaming occurs on SNS. In this research, we aim to develop a system to support the early resolution of flames. This paper report that we tried to classify automatically between pros and cons of opinions using multiple machine learning methods for a tweet thread that burned on Twitter, and obtained results of 86.6-88.1% accuracy.

キーワード: 意思決定支援, 炎上, SNS, ANN, SVM, ロジスティック回帰

Keyword: Decision Making support, Flame, SNS, ANN, SVM, Logistic regression

1. 緒言

近年, Social Network Service以下SNSと呼

ばれるサービスが流行している. SNSは本人の発言や行動に関すること, 事象に対する情報交換などに利用されている反面, 意

見の対立や不適切な発言などで, 多方面から非難や議論が生じてしまう事がある. これを炎上と呼ぶようになった. 炎上には, 刑事事件に発展した事例や悪質なデマ, 過去には炎上によって倒産に追い込まれた事例[1]や名誉が傷つけられ裁判に至る事もある. 裁判となれば, 両者が何で争っているか把握するため, 弁護士は, 各者または各グループの発言を解析し, 要約して, 裁判官に理解してもらえる準備をしないではず負担を要する. 近年は炎上監視するサービス[2]もでており, 早期に炎上を納めることが必要となっているがそのための人員を割くことは企業や個人ではとても対処しきれないので機械を用いることで対処することが最も効率が良い方法だと考えられる.

そこでコンピュータを用いた, 炎上スレッドの解析システムについて研究する.

本ソフトウェア・システムでは, 情報科学の自然言語処理技術を用いて, スレッドの立場毎の発言に分類し, 各発言の要約や特徴を分析する事で, 炎上の解決に有用となるだろう情報の解析, 要約を実施する事を想定している.

本研究では, スレッド内の発言内容から, 発言の立場を特定, 分類する自然言語処理ベースのテキスト・マイニング手法について研究した内容を報告する. 本論文では, 提案システムの設計を示し, 提案する手法の手順や実施結果について報告する.

2. 研究内容

2.1 全体のシステム

提案するシステムは次の通りである.

(ア) 発言を立場に応じたグループ毎へ

分類

(イ) 各グループの発言の特徴量を解析

(i) 感情の解析・可視化

感情を表す単語から感情の強さを元に算出する手法の確立を想定している.

(ii) 信ぴょう性を量る手法の確立

信ぴょう性は, デマへの批判や訂正意見の多さ, 議論の内容の中身が, 感情的か, 知能的かなどで測定できる可能性があり, これらの特徴量を元に算出する手法の確立を想定している. デマでない事を明らかにする事は難しいが, デマに対する反論が多ければ, それは情報の不確かさを示しているといえ, その不確かさが高い情報である事を判別できる事で, デマに扇動される人が減る事を期待できる.

(ウ) 各グループの発言の要約

自然言語処理分野では, 文章の要約の機能が研究されてきており, それらを使う事で要約できる機構をもたせる事を想定している.

(エ) 要約や特徴量に基づく助言の提示

上記で得られた数値を元に, 対応策を助言できる機構をもたせる事を想定している. デマか否かを助言し, 注意を促すなどが実現できると考える.

本論文では(ア)について取り組んだ. その内容は2.2節から述べる.

2.2 AIによる立場の分類の実現

本節では, AIを用い, 発言の立場(例: 賛成派, 反対派, 中立派など)を求める手法を研究する. 一般的に言われるAIは, 機械学

習, 深層学習を使用したものが多く, 本研究では機械学習を用いた手法で研究した.

2.3 提案手法

本研究では, 発言の立場に応じたグループに発言内容をコンピュータで分類する手法を提案する. 機械学習の分類は, 万能ではないが, 主張内容が近いのであれば, グループに分類は可能であると考えた. 以降に, 分類する提案手法の手順について述べる.

発言の類似度の測定を, 本論文では, 発言中の各単語の有無で選別する. そのため, 発言中の単語の有無をベクトルで表現するone-hot-vectorという構造を利用する. One-hot-vectorは, 表 1の通り, 要素1が「私」という単語に割り当てられていたとすると, 「私」という単語が存在する文があれば, 要素1が1となる. 「私」がなければ0が入る.

Bag-of-Word(以下BoW)は, 対象となる文中全ての単語を基に生成したベクトルに対し, 単語の有無でone-hot-vector形式化したものである.

表 1 one-hot-vector

	要素 1	要素 2	要素 3
私	1	0	0
は	0	1	0

2.3.1 手法の手順

Step1 データ解析用の抽出, 整形
激しい議論がなされている一連のテキストデータを解析しやすいファイル形式にスクレイピングする. Webスクレイピングとは, ウェブサイトから情報を抽出するコンピュータソフトウェア技術を指す.

解析用に保存する形式は様々ある. 例えば, 表形式に保存したければTSV, CSV, Excelがある. Excelは特に複数のテーブルを保存できる特徴がある. ExcelやCSVは編集が容易な特徴がある.

また, データの構造(型)が決まっているのならRDBMSなMySQL, PostgreSQLなどがある. また, MySQLの様にサーバの用意がむずかしければSQLiteという単一に利用できる形式がある.

もし, データの構造が決まっていないのならNoSQLに属するMongoDBなどの手段がある. こちらはJSON, BSON, YAMLなどの形式を変換せずDBへ格納する事ができる. この構造は望ましいのは, 近年のSNSが提供しているAPIはJSON形式が多く, サービス提供側のAPIの改修によって提供されるJSONデータの構造も変化する. そのため, 事前にデータ長やデータ型が定められるRDBMSでは変更の負担が大きかった事に起因する. しかし, 現在はMySQLやPostgreSQLなど著名なRDBMSは, JSON型という可変データを格納できる型を提供してくれており, RDBMSが再び着目されるようになっている.

Step2 前処理

前処理では, ノイズ, 悪影響になりえそうな情報の除去やデータ構造の整形を行う. データの除去では, 本手法において意味がない顔文字に使われた記号, 数値, 「を」, 「は」などの助詞などである.

データの整形は, NoSQLのそのまま保存できる事(=非構造のまま保存)と一見反する事項である. しかし, NoSQLは非構造のまま保存できるだけで, 処理する際は, 処理対象を選び, データ型に即して適切に解析す

る必要がある。

Step3 Bag-of-word (BoW) の生成

次に形態素解析で発言内のテキストを単語に分割する。分割して得られた単語を One-shot-vector 形式に落とし込み Bag-of-word を生成する。

Step4 クラス値の付記

生成した BoW のベクトルに、クラス列を追加する。このクラス列は 2.1 節で説明した、機械学習法が、説明変数を基に導出しようとする目的変数を意味する。つまり、目的変数 = クラス列の各セルとなる。このクラス列を、発言が属するグループ (賛成派、否定派など) を One-hot-vector 化で、Group ID へ落とし込む。例えば、賛成派 = 0、否定派 = 1 と定める事を指す。

Step5 弁別モデルの生成

機械学習法を用い、弁別モデルを生成する。

以上で、発言グループを生成するモデルは完成となる。

2.4 実装と評価実験

Step1 で対象の議論を Excel に落とし込む対象を、Twitter のスレッド [3] とした。スレッドとは、つぶやきの連なりを指す。スレッド [3] では、コロナ禍においてマスク着用疑問を持つ娘さんを父が後押しし、マスク着用を奨励しないツイートをした事で、沢山人から反論が寄せられたスレッドである。その賛否両論のつぶやきは現在 209 件となった。本スレッド [3] を Python による Web スクレイピングで、Excel に書き

出した。書き出せた項目を表 2 に示す。

表 2 スレッドより抽出した項目

列名	内容
SC_NAME	表示名
SCREEN_ID	つぶやいた者の ID
DATE	つぶやいた日付
TEXT	つぶやいた内容
REPLY_COUNT	リプライ (返信) の数
RETWEET_COUNT	リツイートの数
FAVORITE_COUNT	Favorite (お気に入り) の数

Step2 に従って前処理を行う。前処理した結果は次の通りである。

- 処理に問題があった NAN (=No ANswer) なレコードは解析から除外した。
- スクレイピング後の text 列には、ツイートの文字列が無いレコードがある。これは、画像のみを添付したツイートであった。

Step3 に従って Bag-of-word を生成する。

形態素解析には janome [4] を利用した。Python で利用できる形態素解析パッケージは多数あるが、janome は Python のみで実装されており、新語や固有語や複合語に強い辞書 ipadic-NEologd も使う事ができる。janome を用いて形態素解析した結果を scikit-learn パッケージの CountVectorize クラスで、BoW を作成した。

Step4 に従い、クラスを表す vector 値を付記する。本事例では、マスク着用の賛成派と反対派である。そこで、各つぶやきにおける発言がマスク着用の賛成派と反対派のどちらの視点で書かれているか、筆者が判断し、Web スクレイピングでスレッド [3] を抽出した Excel へ追記した。この列デー

タがStep4で述べたクラス列に該当する。本Stepでは、そのExcel上の賛成派を0, 反対派を1とone-hot-vector化して、Step3で生成したBoWの各レコードに追記する。

2.5 性能評価とその結果

性能評価の手順は次のとおりである。

Step 1. 訓練データと評価データの分割
訓練データと評価データの使い分けは次のとおりである

1. 訓練データは、議論のグループを識別するモデルを生成するために用いる。
2. 一方、評価データは、生成したモデル

を評価するために用いる。

Step 2. 弁別モデルの生成

Step 1で定めた訓練データに対して機械学習法を用い、弁別モデルを生成する。

Step 3. モデルの性能評価

Step 2で定めた評価データを用い、Step5で生成した弁別モデルの分類性能を評価する。

評価結果は表 3となった。平均スコアが求めるべき性能を表しており、Python コードより得られた数値を少数第 4 位で四捨五入し百分率(%)に変換した。表 3の結果を踏まえた考察を次節で述べる。

表 3 弁別性能スコア

	平均スコア[%]	1 回目	2 回目	3 回目	4 回目	5 回目
ANN	88.1%	92.6%	88.9%	85.2%	85.2%	88.5%
Linear SVM	86.6%	92.6%	88.9%	85.2%	81.5%	84.6%
ロジスティック回帰	88.0%	92.6%	92.6%	85.2%	85.2%	84.6%

2.6 考察

表 3の結果より、本スレッドの分類性能の測定値の平均は高い精度で実現できる事を確認した。約88%の精度は、値的には高い。しかし、実用化のためにはもっと正確性の向上が必要である。その機械学習法もほぼデフォルトのパラメータ設定であるので、パラメータや前処理の最適化を実施していく事で、分類性能を100%近く高められる事が望ましい。精度のさらなる向上には、前処理などを強化していく必要があると考える。

例えば、前処理において、2.4節で述べた通り、・の前処理(Step2)で、text欄が空白の

ツイートを、今回は煽っている内容の画像が多かったので、マスク賛成派に対する煽り画像としてクラス値を全て「賛成派」にした。しかし、削除以外の方法もありそうである。例えば、近年の深層学習を用いる事で、写っている被写体名の抽出や、情景の抽出[5]ができてきているので、画像に写っている内容がわかれば、煽っている内容かそうでないのか、がわかるだろう。煽っているか否かを明らかにできる事と、前後の文脈関係から賛成派、反対派などの立場を明らかにできるかもしれないと考えられる。

3. 結言

SNS上の炎上による訴訟, 広告料獲得のためのデマへの扇動などの問題が生じている. その解決のために両者の意見を解析するAIシステムの開発を目指している. 本論文では, 機械学習法を利用して, 2者の意見を弁別する手法を作成, 評価実験を実施した. 結果, ANNで約88.1%, Linear SVMで約86.6%, ロジスティック回帰で約88.0%の精度で弁別できる事が明らかになった.

今後は, 弁別性能の向上や, 両者の主張の要約をして, ユーザの判断・理解の促進する機能, デマか真実かの信頼度を明らかにする機能, 感情的度合いを可視化する機能などを実装して, ユーザの作業負担や不確かな情報に踊らされないようきっかけを与える意思決定支援システムへと発展させたい.

参考文献

[1] 「『バイトテロ, 一生許せない』あのそば店社長からの手紙」, *NewsPicks*, 12

月 16, 2013.

<https://newspicks.com/news/260500/> (参照 1月 21, 2021).

[2] 「投稿監視サービスの比較! メリット・依頼すべき内容・主要7サービス」, アスピック | クラウドサービス比較・活用サイト.

<https://www.aspicjapan.org/asu/article/969> (参照 1月 23, 2021).

[3] 「いずみchang (久保田 泉)さんはTwitterを使っています」, *Twitter*. https://twitter.com/IZUMI_chang_/status/1306381620337754115 (参照 10月 15, 2020).

[4] T. Uchida, *Janome: Japanese morphological analysis engine*. .

[5] 秀彰柳澤, 拓朗山下と裕渡辺, 「主要キャラクタの抽出を目的とした漫画キャラクタ画像のクラスタリング」, 映像情報メディア学会誌, vol. 73, no. 1, pp. 199- 204, 2019, doi: 10.3169/itej.73.199.