

分散表現を用いた不適切文書判定

平石 健太郎^{1,a)} 柴田 大地¹ 西田 智裕¹ 山口 直子¹ 鈴木 祥太¹ 芳野 魁¹
伊藤 孝行^{1,b)}

概要: 本稿では、投稿文書の分散表現を用いたフィルタリング手法を提案する。近年、オンライン上での議論が活発化している。ただし、これらの議論には無関係のスパムなどの有害なコンテンツが多数あり、また相手を侮辱したり差別したりする激しい発言がある。従って、不適切な発言を削除して安全にオンラインユーザーが参加できる議論環境を構築することが必要になる。不適切な発言を削除するには、文書の意味を理解し分類することが必要である。本稿では、doc2vec を文書のベクトル化、ELMo を単語のベクトル化に用い、ベクトル化された文書を文書類似度計算とディープニューラルネットワーク (DNN) を用いてフィルタを構築した。評価実験では提案手法が高い精度で不適切文書を分類できたことを示す。また実際のリアルタイムの議論でも運用を行い実用性を確認する。

1. はじめに

近年、ウェブ掲示板やソーシャルネットワークサービス (Social Networking Service, SNS) のようにユーザが自由に読み書きのできるコミュニケーションツールが Web 上で非常に発展しており、誰もが Web 上であらゆる情報を発信したり入手することができる。そのためこれらのサイト上では出会い系の書き込み、誹謗中傷する内容、または性的な内容などの閲覧しているユーザを不快にする発言、又は犯罪を助長するような発言が散見され、これらのコンテンツが未成年者にとって悪影響を与えることが考えられる。また、悪影響を与えるコンテンツ (以降有害コンテンツとする) を配信するユーザと未成年のユーザが Web 上を介して接触してしまうことが問題とされており、近年の SNS の成長に伴い、ますます深刻化の一途をたどっている。従って SNS 上および Web 上の有害なコンテンツを判別し、取り除くことで未成年者のユーザに悪影響を与えないようにすることが重要である。

現在、有害コンテンツを取り除く問題として実際に人が目視し、投稿された文書が不適切かどうかを判別する作業を行い、対策をとっている SNS も存在する。しかし、ユーザ数が大規模な SNS も多く、ユーザが書き込む情報も莫大であるため、人間の手作業で不適切な投稿を取り除く作業を行うのは、膨大な時間とコストが掛かってしま

う。従って、近年では不適切な投稿に対して正確に効率よく SNS 上から取り除くことが重要視されてきている。

過去行われた、計算機を用いた有害文書の分類ではベイジアンフィルタを用いたもの [1] や、単語の共起データを用いたもの [2][3] が行われてきた。しかしながら、共起データを用いた分類で精度を向上させるには大規模学習データの処理時間が大きな課題となっている。そのため近年では、深層学習を用いた単語の分散表現やニューラルネットワークを用いた分類が高精度で頻繁に研究されている [8]。

本研究のフィルタリング手法は、伊藤孝行研究室で運営を行なっている「HAMAAgree」でのオンライン議論システムへの実装を想定しているため、処理時間と分類モデルの軽量化が重要になっている。議論中に投稿された、不適切な発言は放置しておく、議論を脱線させたり、炎上を起こす火種になりかねない。従って、議論を阻害する投稿に対しては早急に非表示化する必要がある。

本研究では、まず doc2vec を用いて文書をベクトル化し、コサイン類似度を用いたアンサンブルでの分類手法、および ELMo で単語ごとにベクトル化し、ニューラルネットワークを用いた分類手法について述べる。

評価実験では、提案手法の doc2vec を用いた類似度分類、および ELMo を用いたニューラルネットワークを用いた分類手法を対象に行い、高い評価指標を得られることを確認する。

また、実際の議論でのフィルタリング機能の運用結果について述べ、実用性を確認する。

¹ 名古屋工業大学
〒466-8555, 愛知県名古屋市昭和区御器所町
a) hiraishi.kentaro@itolab.nitech.ac.jp
b) ito.takayuki@nitech.ac.jp

2. 関連研究

2.1 doc2vec

doc2vec (Paragraph Vector)[4] は Le らが提案した文書を分散表現する手法である。doc2vec には Distributed Memory Model of Paragraph Vectors (PV-DM) と Distributed Bag of words version of Paragraph Vector の 2 つのモデルがあるが、本研究では PV-DM を用いた。

図 1 に PV-DM の構造を示す。

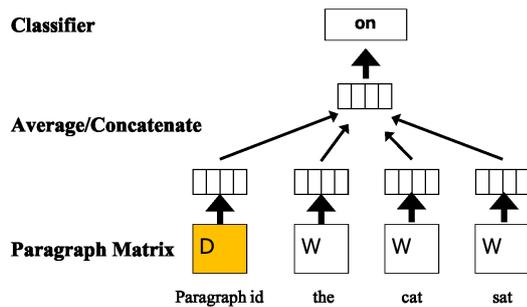


図 1 PV-DM 概要図

文書 i のベクトルは、行列 D における文書ベクトル D_i と表す。PV-DM では文書 i 内の単語 w_t の前 k 単語 w_{t-k}, \dots, w_{t-1} と文書ベクトル D_i から w_t を予測する。

図 1 に示した行列 D , Paragraph Matrix の概要について図 2 に示す。

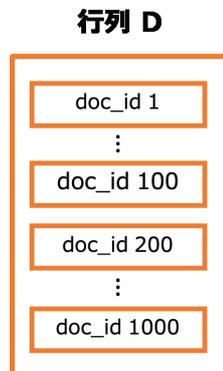


図 2 Paragraph Vector の行列 D の概要

Paragraph Matrix は、それぞれの行が各文書のベクトルに対応している。文書ベクトルのサイズは任意であり、行列の行サイズは学習したい文書の数となる。

PV-DM の最大化したい目的関数は式 (1) となり、対数尤度を最大化する学習を行う。

$$\frac{1}{N} \sum_{t=1}^N \log p(w_t | w_{t-k}, \dots, w_{t-1}, D_i) \quad (1)$$

式 (1) における p は、単語 w_t の前 k 単語 w_{t-k}, \dots, w_{t-1} と文書ベクトル D_i が与えられた時に w_t が出現する尤度で

ある、word2vec の CBOW 同様に、尤度 p はソフトマックスの形で定義され、式 (2) で表される。

$$p(w_t | w_{t-k}, \dots, w_{t-1}, D_i) = \frac{\exp(v_{w_O}^T v_{w_{IN}})}{\sum_{j=1}^V \exp(v_{w_i}^T v_{w_{IN}})} \quad (2)$$

式 2 における w_{IN} は、 w_{t-k}, \dots, w_{t-1} と文書ベクトル D_i の平均であり、 w_O は予測される単語に対応する重みパラメータベクトルである。

2.2 ELMo

ELMo[5] は Matthew らが提案した事前学習によって文脈を考慮した単語表現方法を獲得する手法である、ELMo は文脈を考慮しているため、語彙の曖昧さを解消 (Word Sense Disambiguation) することができる。本研究では ELMo を用いて、各単語の分散表現を獲得した。

3. 文書類似度計算

doc2vec は文書をベクトル化することができるが、ベクトル化された各次元は抽象的な意味を持っている。従って、似ている文書は同じようなベクトルを持っていると考えることができる。提案手法ではラベルづけされた文章をベクトル化して、学習することで各カテゴリの代表ベクトルを作成する。

代表ベクトルと新しい投稿との類似度を計算する手法として、コサイン類似度を用いた。コサイン類似度の式を式 (3) に示す。

$$\cos\theta = \frac{M_i \cdot M_j}{|M_i| |M_j|} \quad (3)$$

式 (3) では、ベクトルが全く同じ方向を向いている場合は 1、逆を向いている場合は -1 を返し、-1 ~ +1 の範囲で値が返ってくる。分類を行うための式を式 (4) に示す。

$$\arg \max_{c \in C} \cos(\vec{q}, \vec{d}_c) = \frac{\sum_{i=1}^{|V|} q_i d_{c_i}}{\sqrt{e \sum_{i=1}^{|V|} q_i^2} \cdot \sqrt{\sum_{i=1}^{|V|} d_{c_i}^2}} \quad (4)$$

C はクラス全体、 c は各分類クラスである。 d_c は各クラスの代表ベクトルであり、 q は新しく投稿された文書のベクトル表現である。式 (4) により各代表ベクトルと新規投稿の類似度を計算し、最もコサイン類似度が高くなったクラスに分類する。

本研究ではアンサンブルによる、分類を提案する。アンサンブル法とは、多数決をとる方法で、個々に別々の学習をさせた分類木を融合させることによって、予測精度を向上させる方法である。

本手法で用いた数式を式 (5) に示す。

$$\arg \max_{c \in C} \frac{1}{N} \sum_{k=1}^N \cos(\vec{q}, \vec{d}_{c_k}) \quad (5)$$

式 (5) で N はアンサンブルするモデルの個数、 d_{c_k} は

各アンサンブルモデルの分類クラスにおける代表ベクトルである。本手法では、doc2vec でモデルを複数個作成し、それぞれのモデルで新しい投稿データと代表ベクトルの間でコサイン類似度を計算し、それぞれ平均化を行い、最も類似度の高いクラスに分類を行った。また、本研究では、アンサンブルに用いるモデルとして、doc2vec でベクトル化する際にベクトル化する次元が異なるモデルを複数用意した。具体的には 300, 600, 800 次元のモデルを利用して行った。図 3 に文書類似度を用いたフィルタリングの流れを示す。

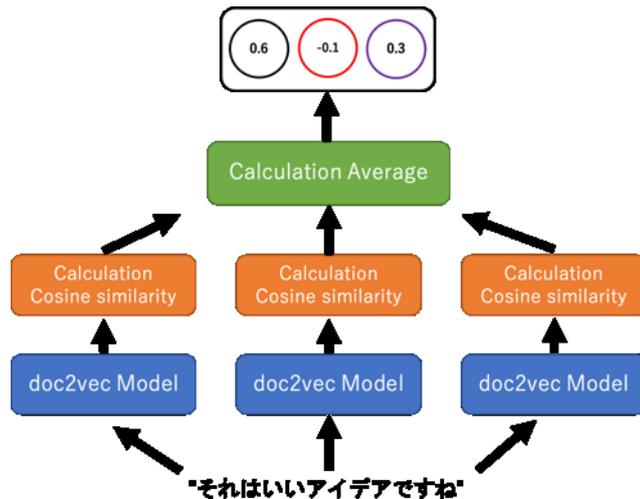


図 3 doc2vec を用いたアンサンブルモデル

4. 情報フィルタの構築

4.1 形態素解析

doc2vec によって文書の分散表現、もしくは ELMo によって単語ごとの分散表現を得るためには、それぞれのモデルで文書を入力する前に対象となる文章を単語単位に分解しておく必要がある。英語などの多くの言語は、単語と単語の間は普通空白によってあらかじめ「分ち書き」されているため、単語単位に分解する必要はない。しかしながら、日本語のような単語を区切るための印がない言語の場合、そのままモデルに入力しても、適切な分散表現が得られない。従って、日本語の文章を適切な形に分解してから、doc2vec や ELMo などのモデルに入力する必要がある。言語の種類によってフィルタ制度に差異が発生する。上記のような課題に対応するために形態素解析 (Morphological Analysis) などの自然言語処理を追加する必要がある。

形態素解析とは、コンピュータ等の計算機を用いた自然言語処理技術の 1 つであり、かな漢字変換等にも応用されている。文法の知識 (文法の規則の集まり) や辞書 (品詞等の情報付きの単語リスト) を情報源として用い、自然言語で書かれた分を形態素 (Morpheme, 言語の最小単位, 大まかに意味をもつ) の列に分解し、それぞれの品詞を判

別する作業を指す。

本研究では MeCab[6] という文を形態素に解析するツールを使用した。Mecab を用いて「すもももももものうち」を形態素解析した例を図 4 に示す。

すもももももものうち	
すもも	名詞, 一般, *, *, *, *, すもも, スモモ, スモモ
も	助詞, 係助詞, *, *, *, *, も, も, も
もも	名詞, 一般, *, *, *, *, もも, もも, もも
も	助詞, 係助詞, *, *, *, *, も, も, も
もも	名詞, 一般, *, *, *, *, もも, もも, もも
の	助詞, 連体化, *, *, *, *, の, の, の
うち	名詞, 非自立, 副詞可能, *, *, *, うち, ウチ, ウチ
EOS	

図 4 MeCab の解析結果

また、形態素解析の際、それだけで意味のなさないもの、また記号なども不適切な発言には多く見られることから、全ての形態素を学習対象とした。

4.2 回帰結合型ニューラルネットワーク

回帰結合型ニューラルネットワーク (Recurrent Neural Network, RNN) はテキストデータなどの入力データの順序によって出力が変わる場合に有効である。

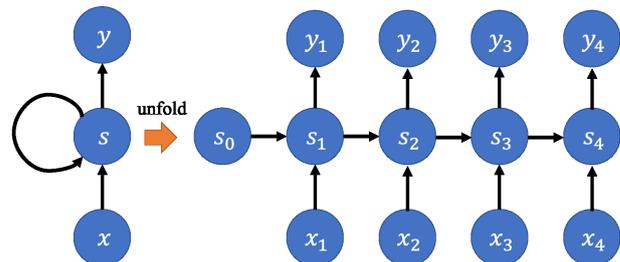


図 5 RNN の概要

図 5 の左側はループを持たせた回路図で、右側は同じネットワークを展開した計算グラフであり、各ノードはそれぞれ 1 つの時間における値と関連づけられている。ある時点 t での入力値 x_t と前回時点での状態 s_{t-1} から新しい状態 s_t に状態が遷移する。そして状態 s_t から出力値の y_t が出力される。RNN は式 (6) で定義することができる。

$$\begin{aligned}
 y_t &= f(s_t; \theta) \\
 s_t &= h(s_{t-1}, x_t; \theta)
 \end{aligned}
 \tag{6}$$

RNN は入力データの記憶を保持させるために、中間層にループを持たせている。従って、ニューラルネットは 1 個前のデータを判断材料に使うことができるため、当然 1 つ前のデータを扱うときはそのまた 1 つ前のデータとの依存性を扱うことができるはずである。ただ実際は学習がうまくできないなどの問題であまり使われることはなかった。

Long Short-term Memory (LSTM) ネットワーク [7] は RNN を拡張し、前の情報をうまく扱うことに特化し、長期的な依存関係を学習できるものである。図 6 に LSTM の構造を示す。

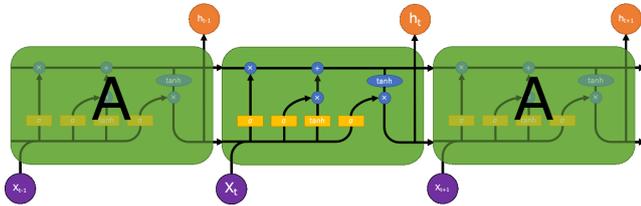


図 6 LSTM の概要

LSTM での各ゲートでの情報の取捨選択は、シグモイド関数で行われる。出力が 0 であれば、ゲートを通さず、1 であれば全てを通す。

また Gated Recurrent Unit (GRU) は、LSTM を少しシンプルにしたモデルであり、LSTM での入力ゲートと忘却ゲートを越すインゲートとして 1 つのゲートに統合している。図 7 に GRU の構造を示す。

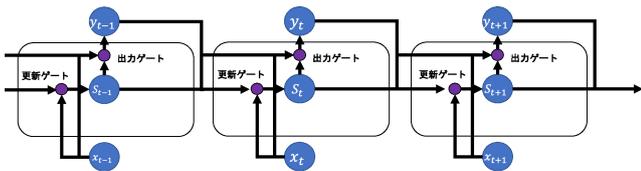


図 7 GRU の概要

LSTM と同様に忘却・更新ゲートを導入することにより、各時間ステップ間で迂回するショートカットパスが効率的に生成することができ、長いステップ前の出来事の特徴の記憶を維持しやすくなることができる。従って学習中に誤差を容易に逆伝播することができ、勾配消失の問題を解決することができる。

LSTM や GRU は通常、入力順すなわち前向きに依存関係を学習するが、本研究では、双方向 (Bidirectional) に学習を行うことができる、Bi-LSTM, Bi-GRU を用いた。

4.3 フィルタリングの構築手法

フィルタリングの構築は以下のステップ 1, ステップ 2, ステップ 3, およびステップ 4, もしくはステップ 5 の流れで作った。(図 8)

ステップ 1

学習データは Web 上からクローラーを用いて学習データを自動で収集し、収集に用いるサイトは 5 ちゃんねる掲示板、および本研究室で運用を行っている COLLAGREE, D-Agree とした。今回の実験では約 1 万件のデータを収集した。

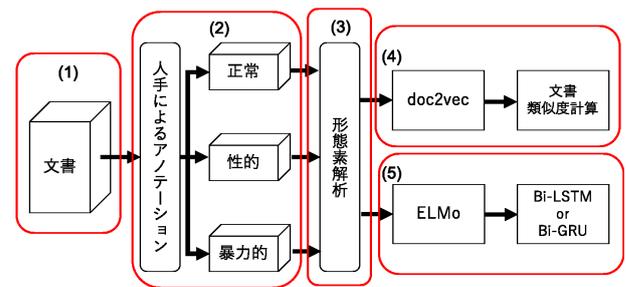


図 8 フィルタリングの構築

ステップ 2

収集した学習データを正常な投稿、性的 (卑猥な発言や、出会い系の発言など) な投稿、暴力的 (攻撃的、差別的、侮辱的など) な投稿に分類する。ラベル付けは、人の目視により行った。

ステップ 3

形態素解析を用いて学習データを単語分割する。本研究では、MeCab を形態素解析器に用いて、ステップ 4 もしくはステップ 5 に入力した。

ステップ 4

形態素ごとに解析された文書データを doc2vec モデルに入力し、分散表現にする。フィルタの構築では分散表現になった各ラベルの文書を学習し、代表ベクトルとする。生成された代表ベクトルと新規投稿の間で類似度計算を行い、最も値が高くなったクラスに分類する。

ステップ 5

形態素ごとに分割された文書を形態素ごとに ELMo で分散表現にする。分散表現となったものを Bi-LSTM, Bi-GRU のレイヤーに入力として、それぞれの出力を出力層に渡す。出力層の活性化関数は softmax を用いて、各クラスの確率を求め、最も確率の高いクラスに分類する。

5. 評価実験

5.1 評価指標

本研究では、文書分類手法の評価指標として、テストデータから不適切文書を分類する際の評価結果の指標として F 値を用いる。F 値は適合率と再現率の調和平均である。分類手法の性能は主に正確性と網羅性の質的な観点から適合率 (Precision) と再現率 (Recall) を用いる。適合率と再現率を分類性能の良的な観点から測定することにより判定する。式 (7) に F 値の式を示す。

$$F - measure = \frac{2Recall \cdot Precision}{Recall + Precision} \quad (7)$$

適合率とは、文書分類結果として得られた集合の中にどれだけ正しく分類できた文書を含んでいるかという正確性の指標である。また、再現率は分類対象としている文書の中で分類結果として適合している文書のうち、どれだけの文書を分類できているかという網羅性の指標である。言い換えると適合率は、「暴力的」と分類された文書のうち、ど

れだけ正しく分類されているかという”正確性”を表し、再現率は、テストデータに含まれる暴力的文書をどれだけ「暴力的」と分類されているかという網羅性を表す。例えば、暴力的文書の適合率が低く、再現率が高い場合は、テストデータの暴力的文書を正しく分類していても、同時に多くの正常、性的文書も「暴力的」と分類してしまっていることになる。以上の理由より、本研究では各クラス分類における F 値の平均を採用する。

5.2 実験設定

評価実験では、各分類手法の分類精度を示すために、実際のコミュニティサイトからデータを収集し、学習を行い、同様に収集したデータをテストデータとして分類する。

学習データ

本実験の学習データとして、本研究室で運用を行っている COLLAGREE[12]、および D-Agree 上の投稿データ、また大規模オンライン掲示板である 5ちゃんねる掲示板からデータを収集した。正常発言を 3,200 件、性的発言を 2,400 件、暴力的発言を 1,920 件、合計 7,520 件の投稿文書を用いて、学習データの生成を行う。

テストデータ

テストデータは学習データと同様に COLLAGREE, D-Agree および 5ちゃんねる掲示板から収集した正常発言 800 件、性的発言 600 件、暴力的発言 480 件を用いて行う。また、学習データとは異なる文書を用いる。

5.3 実験結果

まず、それぞれの分類クラスにおける F 値の平均を表 1 に示す。

表 1 実験結果 (F 値平均)

Method	F-measure
doc2vec Ensemble	0.936
Bi-LSTM	0.919
Bi-GRU	0.916

次に各カテゴリにおける分類結果の詳細について示す。表 2 は正常発言文書クラスにおける分類結果を示し、表 3 は性的発言文書は性的文書クラスにおける分類結果、表 4 に暴力的発言文書クラスにおける分類結果を示す。

表 2 正常発言文書の実験結果の詳細

Method	Precision	Recall	F-measure
doc2vec Ensemble	0.931	0.988	0.959
Bi-LSTM	0.928	0.945	0.937
Bi-GRU	0.902	0.970	0.935

表 3 性的発言文書の実験結果の詳細

Method	Precision	Recall	F-measure
doc2vec Ensemble	0.960	0.904	0.931
Bi-LSTM	0.934	0.939	0.936
Bi-GRU	0.954	0.914	0.934

表 4 暴力的発言文書の実験結果の詳細

Method	Precision	Recall	F-measure
doc2vec Ensemble	0.931	0.906	0.918
Bi-LSTM	0.901	0.867	0.884
Bi-GRU	0.918	0.846	0.880

フィルタリング機能は、表示すべきか、表示しないべきかの 2 パターンとなるため、正常発言を無害 (None) とし、性的、および暴力的発言を有害 (Toxic) とし、無害か、有害かどうかの 2 値分類を行った結果を表 5 に示す。表に示す値は全て F 値である。

表 5 2 クラス分類結果
(値は全て F 値)

Method	None	Toxic	Average
doc2vec Ensemble	0.959	0.966	0.962
Bi-LSTM	0.937	0.951	0.944
Bi-GRU	0.935	0.947	0.941

各クラスにおいて、doc2vec を用いたアンサンブルモデルによる文書類似度分類が最も高い精度を示した。

5.4 先行研究との比較

先行研究との比較として、過去行われてきた単語共起フィルタリング、および Paragraph Veector と多層パーセプトロンを用いた有害文書判定との精度比較を行う。過去行われてきたフィルタリング手法は 2 クラス分類であるため、2 クラス分類の精度を以下の表 6 示す。

表 6 先行研究との比較

Method	F-measure
doc2vec Ensemble (propose)	0.962
Bi-LSTM (propose)	0.944
Bi-GRU (propose)	0.941
PV-CBOW [8]	0.943
Naive Bayes - Bayesian Filter [9]	0.93
Gray Robinson - Bayesian Filter [10]	0.884

先行研究とはデータセットや、データ数が異なるため一概に評価はできないが、提案手法は先行研究の手法より高い F 値を得られることがわかった。

6. リアルタイムでのフィルタリング運用

本フィルタリング手法は 2018 年 11 月 1 日から 2018 年 12 月 7 日まで行われた名古屋市次期総合計画案についての社会実験に用いられたシステム「HAMAAgree」に実装を行った。

6.1 HAMAgree

HAMAgree は本研究室で、開発・運用を行っているオンラインにおける議論支援システムである。HAMAgree は多くの参加者の間での議論を支援する自動ファシリテーターが実装されている。

オンライン議論におけるファシリテーター [17][18] とは合意形成のために議論を誘導、統合、要約を行い、議論を建設的に進行させる役割である。



図 9 HAMAgree のトップ画面

6.2 HAMAgree におけるフィルタリング

HAMAgree での炎上、有害な発言、および参加ユーザの誹謗中傷を防ぐために doc2vec を用いたアンサンブルによるフィルタリングを実装した。フィルタリング機能は先に述べたエージェントのうちのマネージャーエージェントに実装を行った。ここで doc2vec を用いた理由として、精度が最も高かったことに加え、計算時間が非常に短いこと、またモデルのサイズが非常に軽量でマネージャーエージェントに実装が容易であるためである。本フィルタリング機能は、投稿を削除するわけではなく、一時的に非表示にし、非表示になった投稿を再度人の目でチェックし、不適切でなければ再度表示、不適切であれば削除という方法で行っている。

炎上しそうな投稿

図 10 に本フィルタが不適切と判断した投稿を示す。



図 10 炎上しそうな投稿

図 10 で示している p,n,v はそれぞれ正常発言 (Positive)、性的発言 (Negative)、暴力的発言 (Violence) であり、数値は各クラスの代表ベクトルとのコサイン類似度の値となっている。この投稿は不適切な発言ではないが、長文であり、読み手にとってはあまり肯定的な発言とは取れず、誤解を招きかねず、今後の議論が炎上に向かう可能性がある。本フィルタでは、この投稿に対して暴力的ベクトルとのコサイン類似度が最も高かったため、暴力的と判断し、この投稿を一般ユーザには表示されない非表示状態とした。

砕けた投稿



図 11 砕けた投稿

図 11 も図 10 と同様に本フィルタが不適切と判断した投稿である。

図 11 で不適切と判断された投稿は共に「w」や「(笑)」が含まれており、砕けた感じの投稿となっている。ただ投稿としては、閲覧者を不快にするわけではないためあまり不適切な発言ではない。しかし、この投稿に対して、本フィルタは性的と判断してしまっている。

以上のように、炎上が起こりそうな発言に対して本フィルタは機能した反面、記号やネットスラングなどに対して過剰に反応し、閲覧者を不快にさせたり、議論を妨害しているわけではない投稿に対しても、不適切と判断してしまっ

7. まとめと課題

F 値を指標とした評価実験を行った結果、doc2vec を用いたアンサンブルモデルの文書類似度分類手法が最も高い F 値を示した。ELMo で形態素ごとに分散表現にし、Bi-LSTM および Bi-GRU のニューラルネットモデルよりも約 1%F 値が高くなった。doc2vec を用いたフィルタリングは非常に計算を高速に行え、実際のオンライン議論でも誤解を招きかねず、炎上を起こしかねない投稿に対してもタイムレスに不適切と判断し、炎上を防ぐことができた。

しかしながら、ネットスラングや記号などを用いた顔文字などの砕けた投稿に対して過剰に反応してしまい不適切

としてしまうなどの問題点もあった。ネットスラングなどの用語が日常的に使われている面もあるため、このような単語への対応は非常に重要である。そのため学習データにこのような砕けた発言を正例として含めて学習することで解消できると考える。

また、本研究では不適切な投稿のカテゴリとして、性的及び暴力的の2種類に分類を行ったが、不適切な文書はさらに細分化することができ、それに対応したさらなるデータが必要となると考える。

謝辞

研究内容は、JST CREST「エージェント技術に基づく大規模合意形成支援システムの創成：代表伊藤孝行」（グラント番号 JPMJCR15E1）に支援を受けている研究の一部である

参考文献

- [1] Paul Graham: Better Bayesian Filtering; *Proceedings of the 2003 Spam Conference*, (2003)
- [2] Satoshi Ando, Yutaro Fujii, Takayuki Ito: Filtering Harmful Sentences based on Multiple Word Co-occurrence; *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*, (2010)
- [3] Yutaro Fujii, Takuya Yoshimura, Takayuki Ito, Satoshi Ando: Filtering Harmful Sentences based on Three Words Co-occurrence; *In the Proceedings of The 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS 2011)*, (2011)
- [4] Quoc V. Le, Tomas Mikolov: Distributed Representations of Sentences and Documents; *JMLR W&CP 32 (1)*, pp.1188-1196 (2014)
- [5] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer: Deep contextualized word representations; *NAAACL 2018*, (2018)
- [6] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis; *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp.230-237, (2004)
- [7] Hochreiter, S. and Schmidhuber, J: Long Short-Term Memory; *Neural Computation*, Vol.9, pp.1735-1780 (1997)
- [8] 佐藤元紀, 伊藤孝行: Paragraph Vector を用いた有害文書分類手法; 合同エージェントワークショップ&シンポジウム (JAWS2014), pp.321-324, (2014)
- [9] 大塚孝信, Deyue Deng, 伊藤孝行: 3単語共起フィルタリングによる有害文書分類手法と大規模データ処理; 電気学会論文誌 134.1, pp.168-175 (2014)
- [10] 吉村卓也, 藤井雄太郎, 伊藤孝行: Paul Graham 及び Gray Robinson に基づく単語共起フィルタの検証; 第10回情報科学技術フォーラム (FIT2011), (2011)
- [11] Takayuki Ito, Yuma Imi, Takanori Ito, and Eizo Hideshima: COLLAGREE: A facilitator-mediated large-scale consensus support system; *Collective Intelligence 2014*, (2014)
- [12] Takayuki Ito, et al.: Incentive Mechanism for Managing Large-Scale Internet-Based Discussions on COLLAGREE; *Collective Intelligence 2015*, (2015)
- [13] Akihisa Sengoku, Takayuki Ito, Kazumasa Takahashi, Shun Shiramatsu, Takanori Ito, Eizo Hideshima, and Katsuhide Fujita: Discussion tree for managing large-scale internet-based discussions; *Collective Intelligence 2016*, (2016)
- [14] 伊藤孝行: エージェント技術に基づく大規模合意形成支援システムの創成; 経営システム, 経営工学会, (2018)
- [15] 西田智裕, 伊藤孝行, 福島大地, 仙石晃久, 伊藤孝行: ワークショップにおける対面式と非対面式を組合せた手法の検証; 日本デザイン学会研究論文集, 64巻4号, (2017)
- [16] 伊藤孝行, 深町駿平, 杉山弓香, 西田智裕, 秀島栄三, 伊藤孝行: 合意形成支援システムを利用した域学連携手法の有効性; 日本建築学会計画系論文集, 第82巻, 第742号, (2017)
- [17] 伊美裕麻, 伊藤孝行, 伊藤孝行, 秀島栄三: オンラインファシリテーション支援機構に基づく大規模意見集約システム COLLAGREE - 名古屋次期総合計画のための市民議論に向けた社会実装; 情報処理学会論文誌, (2015)
- [18] Takayuki Ito: Towards Agent-based Large-scale Decision Support System: The Effect of Facilitator; *In The 51st Hawaii International Conference on System Sciences (HICSS2018)*, (2018)
- [19] 伊藤孝行, 奥村命, 伊藤孝行, 秀島栄三: 多人数ワークショップのための意見集約支援システム Collagree の試作と評価実験～議論プロセスの弱い構造化による意見集約支援～; 日本経営工学会論文誌, Vol.66, No.2, (2015)
- [20] Noble, Douglas and Rittel, Horst W.J.: Issue-Based Information Systems for Design; *Computing in Design Education*, pp. 275-286, (1988)