

感情パーソナリティの可視化による Twitter ユーザの探索および評価支援

Visualization of Emotional Personality to Support Exploration and Evaluation of Twitter Users

工藤 裕登^{1*} 砂山 渡² 服部 峻³
Yuto Kudo¹ Wataru Sunayama² Shun Hattori³

¹ 滋賀県立大学大学院 工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

^{2 3} 滋賀県立大学 工学部

^{2 3} School of Engineering, The University of Shiga Prefecture

Abstract:

Nowadays, SNSs are used to communicating with friends and gather information. Therefore, users often search for new friends or for users who have the information they want to know when they join a community. In this study, we propose a system that supports the search and evaluation of potential users with whom users can interact and gather information by inputting keywords related to the users to be searched for, using Twitter as a subject. The proposed system extracts and visualizes emotional personalities based on the frequency of use and trends of emotional words in the tweets of users collected by the system. Experimental results show that the proposed system is effective in supporting user search and evaluation.

1 はじめに

近年, SNS の利用者が増加しており, さまざまなユーザとの交流が非常に盛んに行われている. SNS は多種多様なユーザとの交流するためだけでなく, 趣味や仕事などの情報収集として用いられている. しかし多くのユーザの発信内容を確認し, 欲しい情報を探索することやユーザとの交流を実際に行うか判断することは非常に大きな手間がかかってしまう. 自分と相性の良い相手や SNS の内容を読んでいて不快にならない相手を探すために, ユーザの「感情」の表現方法に着目することが重要であると考え.

そこで本研究では, SNS の一つである Twitter[1] を題材とし, ユーザの Twitter の投稿内容 (以下, 「ツイート」とする) に含まれる感情を表現している単語やツイートの使用回数や使用傾向を算出し可視化することで, システム利用者の重視する感情を考慮したユーザの探索および評価を支援することを目的としている. また, 可視化するユーザの感情表現の特性を「感情パーソナリティ」と定義する.

2 関連研究

本章では, Twitter を題材とした感情表現の分析またはユーザ推薦についての関連研究について述べる.

近年の社会経済に影響を与えた新型コロナウイルス (以下, 「コロナ」とする) に関するツイートを対象として, 主要な感情の数値化を行う研究 [2] がある. コロナ関連のツイートにおける「嫌気」の感情の推移と人出との関係を検証し, ネガティブな感情値が高まると約 1ヶ月後に人手が減少する傾向を確認している. また, 韓国で発生した中東呼吸器症候群 (以下, MERS) 状態での韓国版 Twitter における感情分析を行う研究 [3] がある. 感情辞書と言語固有の特徴量を機械学習で分析し, 中立なツイートと感情を持つツイートに分類する手法を提案している. しかし, これらの研究では, 複数のユーザによる感情の分析や感情の推定を行っているが, ユーザー一人一人のツイート内容までは分析していない. 本研究では, 収集した複数ユーザから感情表現を用いて探索したいユーザを抽出し, 実際の感情表現の使用方法を確認しながら評価を行うことを支援している.

また, ユーザ間のフォロー (ユーザのツイートを継続的に表示させること) 関係と実際のやりとりの有無を組み合わせ, 新たな手法を提案している研究 [4] が

*連絡先: 滋賀県立大学大学院工学研究科 電子システム工学専攻
工藤 裕登
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: ov23ykudo@ec.usp.ac.jp

ある．やりとりの関係の有無による従来手法と比較した推薦するユーザの比較を行っている．また，Twitter内の感情を可視化しユーザを推薦する研究 [5] がある．ツイート内の単語に感情を付与し，単語に対して類似する感情を持つユーザを推薦する研究である．これらの研究では，実際に使用される感情表現を確認しながら，ユーザを探索し評価するシステムにはなっていない．そこで本研究では，感情表現を組み合わせで探索したいユーザを抽出し，実際のツイート内容を確認しながら評価を行うことを支援するシステムを提案する．

3 本研究で扱う感情表現

本章では，本研究で扱う感情表現について述べる．

表 1: 本研究で用いた感情表現単語辞書と単語例

感情表現	ポジティブ, ネガティブ分類	単語総数	単語例
「喜」	ポジティブ	143	嬉しい, 優しい
「楽」	ポジティブ	214	楽しい, 笑い
「愛」	ポジティブ	235	恋, 好き
「驚」	ポジティブ	52	仰天, 久々
「怒」	ネガティブ	113	激昂, 叱る
「哀」	ネガティブ	326	悲しい, 悔しむ
「憎」	ネガティブ	383	嫌い, 悪口
「恐」	ネガティブ	277	怖い, 戸惑う

人間の持つ感情に対して，中村 [6] の感情表現辞典では，実際の表現をもとに「喜」，「怒」，「哀」，「怖」，「恥」，「好」，「厭」，「昂」，「安」，「驚」の 10 の感情を定義している．また，人間の持つ「喜」，「怒」，「哀」，「楽」，「愛」，「憎」の 6 つの感情を「六情」とする説も古来から存在する感情分類の一つである [7]，[8]．

そこで本研究では分析の対象とする感情について，中村の定義した 10 の感情や六情をもとに 8 つの感情を定義した．また，これらの感情を表現している単語を「感情表現単語」と定義し，感情表現単語辞書の作成を行った．感情表現単語辞書には，感情の定義をもとに感情表現辞典や日本語感情表現辞書 [9] を用いて感情表現単語の収集を行った．

さらに，8 つの感情表現をポジティブ，ネガティブに分類することで，感情表現単語を含むツイートを「感情ツイート」と定義する．感情表現単語と同様に，ポジティブツイートやネガティブツイートに分けることによって処理の対象とした．

本研究で用いた感情表現単語辞書と単語例を表 1 に示す．入力したユーザのツイート内に含まれる感情表現単語と感情ツイートをを用いて，ユーザの感情パーソナリティの算出を行う．感情表現単語の使用回数および使用割合や感情ツイートのツイートおよび偏差値を，

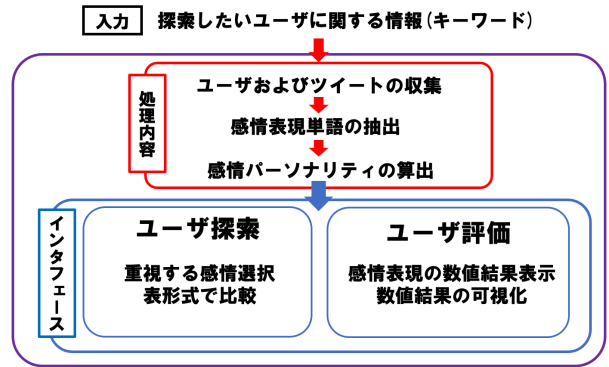


図 1: システムの流れ

感情パーソナリティとして可視化を行う．可視化した感情表現の情報を用いて，ユーザの探索および評価を支援する．

4 感情パーソナリティの可視化による Twitter ユーザの探索および評価支援システム

本章では，感情パーソナリティの可視化による Twitter ユーザの探索および評価支援システムについて述べる．本システムはテキストデータのための総合環境 TETDM(Total Environment for Text Data Mining)[10] 上に表示される画面(以下，「パネル」とする)に処理結果を出力し構成する．図 1 には，システムの流れについて示している．

4.1 探索したいユーザに関する情報の入力

本研究では，Twitter ユーザとツイート内容の収集に Twitter Search API[11] を利用した．探索したいユーザに関する情報の入力として，「キーワード」を入力する．キーワードについて発信しているユーザを収集し，収集した各ユーザに対して 1 ユーザあたり最大 100 ツイートの収集を行う．

4.2 感情表現単語および感情ツイートの抽出

1 ユーザのツイート内の各単語について，感情表現単語辞書に含まれる単語が出力された時，各感情表現に分類を行う．同時に，抽出した感情ツイートをポジティブ，ネガティブツイートに分類する．全てのツイートに対して抽出を行い，感情表現単語の単語数と感情ツイートのツイート数をカウントし感情パーソナリティ

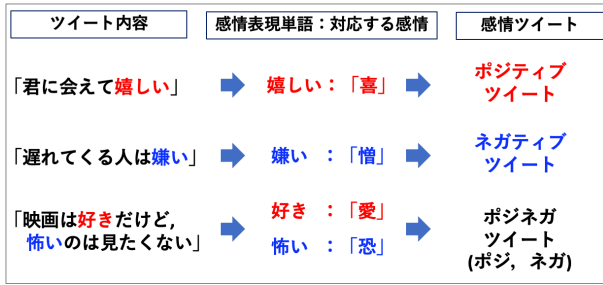


図 2: 感情表現単語および感情ツイートの抽出

の算出に用いる。図 2 に感情表現単語および感情ツイートの抽出方法を示す。

4.3 感情パーソナリティの算出

感情パーソナリティの算出には、8つの感情表現単語の使用回数および使用割合やポジティブ、ネガティブに分類した感情ツイートのツイート数と偏差値を用いる。感情表現単語の使用割合の計算を数式 (1) に、感情ツイートの各偏差値に用いるツイート数および感情表現の傾向を表 2 に示す。使用割合の算出は、各感情表現単語の使用回数と全ての感情表現単語の使用回数をを用いる。偏差値については、無作為に抽出した 5000 ユーザの感情ツイート、ポジティブツイート、ネガティブツイートに対してもツイート数を用いて偏差値を算出する。算出した偏差値を用いて、感情表現の傾向を 3 段階に分類し、感情表現やポジティブ感情、ネガティブ感情を表現する傾向を算出する。

$$\text{各感情の使用割合 (\%)} = \frac{\text{抽出した各感情の単語数}}{\text{抽出した全感情単語数}} \times 100 \quad (1)$$

表 2: 各偏差値に用いるツイート数および感情表現の傾向

偏差値	60 以上	40～59	40 未満
感情ツイート	60.3 以上	25.5～60.2	25.5 未満
ポジティブツイート	50.7 以上	17.1～50.6	17.1 未満
ネガティブツイート	25.1 以上	4.9～25.0	4.9 未満
感情表現の傾向	高い	普通	低い

感情表現単語を対象とする「単語ベース」の処理結果と、感情ツイートを対象とする「ツイートベース」の処理結果に分けて、ユーザの感情パーソナリティの算出する。算出した単語ベースおよびツイートベースの結果について、感情パーソナリティとして可視化を行うことによって、ユーザの探索および評価に対して支援を行う。

4.4 ユーザ探索支援インタフェース

複数ユーザから探索したいユーザを抽出するユーザ探索支援インタフェースを図 3 に示す。探索したいユーザに関するキーワードを入力し、複数のユーザとユーザのツイートを収集する。また、キーワードを入力することで、すでに収集している複数ユーザを再度表示させるパネルも設定している。収集したユーザについて、利用者がユーザに重視する感情 (またはキーワード) を選択することで、表形式表示内に選択結果を反映することができる。選択結果には、選択パターンに該当する重みとツイートの偏差値の算出に用いた 5000 ユーザの感情表現の使用回数をもとに算出した評価点を用いて、各感情およびキーワードに対する重みと評価点の積の総和を反映する。選択結果や使用したキーワードの合計などを用いて表形式表示のソートを行うことで、探索したいユーザに重視する感情表現やキーワードを考慮したユーザの抽出を行うことができる。抽出したユーザの感情表現単語のカラー表示と感情パーソナリティの簡易表示を確認することで、抽出したユーザがどのような感情表現単語や感情ツイートをを用いているかを素早く確認することができ、探索したいユーザ候補であるか判断できる。探索したいユーザ候補を抽出した後に、ユーザ候補に対して感情パーソナリティの簡易表示内に設定しているボタン 1 つでユーザ評価インタフェースに切り替えることができ、素早く抽出したユーザの感情パーソナリティの可視化表示内容を詳しく分析することができる。

4.5 ユーザ評価支援インタフェース

抽出したユーザ候補の感情パーソナリティの可視化表示内容の詳細を確認するユーザ評価支援インタフェースを図 4 に示す。ユーザ探索支援インタフェースで用いた感情表現単語のカラー表示と感情パーソナリティの簡易表示に加え、単語ベースとツイートベースの数値結果や使用割合のレーダーチャート、ツイートに含まれる感情表現の分布、感情ツイートの使用傾向の可視化を用いてユーザの評価を行う。単語ベースとツイートベースの数値結果を確認し、単語ベースやツイートベースで着目したい感情表現を確認する。単語ベースの可視化表示内容では感情表現単語の使用割合をレーダーチャートを用いて可視化し、感情表現単語をどの程度使用しているかを視覚から得ることができる。また、ツイートベースの可視化表示内容において、感情ツイートの使用傾向の可視化を見ることで、抽出したユーザの感情表現の傾向を把握する。さらに、ツイートに含まれる各感情の分布の可視化によって、どのツイートにどの感情表現が用いられるかを視覚から得ることができ、感情表現を含むツイートや含まないツイート

キーワード入力

検索するキーワードを入力してください

キーワードは「word1、word2」で区切ることができる

このとき、word1とword2を含むツイートを投稿しているユーザを集めます

被験者名

キーワードファイル名

大学、サークル

感情表現単語のカラー表示

大学演劇サークルなどに行きたいけど行ける人かな

発作って書いても女王とナレターってまず声変えないや？だよな

うちの部員の課題提出やばいっ？

深夜に文化祭の動画見るもんじゃないねえん

1年Zは着替えを持ってこないし予想今日私の世間の職員室覗きに行ったらからみんなは見たとんや

明日は先生に話しかけてあげおめしな

年賀状しるまでしてもらお3時間分の傑作2つの飯ですに放牧ヨグル(私乳製品食べない飯屋おこす)が当たってるんだけど怖くない？

合衆国(海外)なるあ

学年末の次の日セカオワライブで休むからみんな日部送ってねお願いして私は静岡ですてへ

夏後のからの地域からの学年末からのセカオワライブで分が分解するさるけど頑張れるモチベ

私まじでするってしたらイラストに描ける

お夢きちゃんからの年賀状の板図好き

頭切すぎで先生と早く話したい欲しが高まらん勉強してくれ

私の親い痴でって素でことだと思っ

ことし夏休まで家で勉強の文系科目かんばる体重増やす 身長伸ばす先生と沢山出する→好きな女になる今年もセカオワ1色にする機嫌悪いしないギターでセカオワ弾けるようにする量産はお菓子作りですって食べるレベルにする

妹がが1年何するかもとめてたけど波紋作る気力がようやく出来たことに涙が出そうで出ませんでした

もう好きな人と会えた！！！！！うん

ゆまじたら離婚

感情パーソナリティの簡易表示

テキストユーザ名: ****

ツイート数: 100ツイート

感情を含むツイートの偏差値傾向

ポジネガツイートの偏差値より

ポジティブな感情を表現する傾向

ネガティブな感情を表現する傾向

8つの感情表現より

最もよく使用する感情表現

最もよく使用された感情表現単語

ユーザ比較表形式表示

ユーザ比較

リセット

重視する感情選択

選択肢 特に重視 重視 少し重視どちら... 重視し...

喜

怒

哀

愛

憎

驚

恐

ポジ

ネガ

大学

サー

図 3: ユーザ探索支援インタフェース

をピックアップして表示することが可能になる。これらの機能を用いてユーザ評価を行った後に、再度ボタン1つでユーザ探索に戻りユーザ候補を抽出することで、ユーザの探索および評価を繰り返し行うことが可能になる。これらのユーザ探索および評価を感情パーソナリティを用いて支援する。

5 感情パーソナリティの可視化による Twitter ユーザの探索および評価支援の検証実験

12名の被験者に、4つの課題を用いてキーワードを使用して収集した約100名のユーザから目的に応じたユーザの探索および評価を行ってもらった。ユーザ探索において課題に用いる重視する感情選択パネルの有無(図5)によって被験者を6名ずつのグループに分けて実験を行った。ユーザ探索で5名のユーザを抽出し、ユーザ評価でフォローしたいユーザを判断してもらった。

キーワードの入力とインタフェースの切り替えについては被験者は操作を行わずに実験を行った。

5.1 実験設定

表3に提案システムの検証実験課題を示す. また, 各課題における目的の内容について次に示す. 課題3, 4

表 3: 提案システムの検証実験課題

課題番号	課題 1	課題 2	課題 3	課題 4
キーワード	海外 旅行	就活 新卒	英語 勉強	コロナ 対策
目的	友達探し	友達探し	情報収集	情報収集
グループ 1 感情選択パネル	なし	あり	なし	あり
グループ 2 感情選択パネル	あり	なし	あり	なし

の「情報収集」には課題1, 2の「友達探し」に類似するキーワードを設定し、目的に応じたユーザの探索および評価についての違いについても考慮している。

- 友達探し：お互いにコミュニケーションを取る
- 情報収集：一方的に情報を得る

実験で使用するインタフェースについて、実験用ユーザ探索インタフェースを図5に示し、実験用ユーザ評価インタフェースは図4で作成したものを共通して用いる。実験用ユーザ探索インタフェースでは重視する感情選択パネルの有無により、被験者の好みを反映したユーザ探索および評価の抽出方法の違いや実際にフォローしたいユーザ数などについて比較を行った。感情表現単語数やキーワード合計などでの単純なソートでの抽出結果と好みを反映した選択結果によるソートでの抽出結果について比較する。

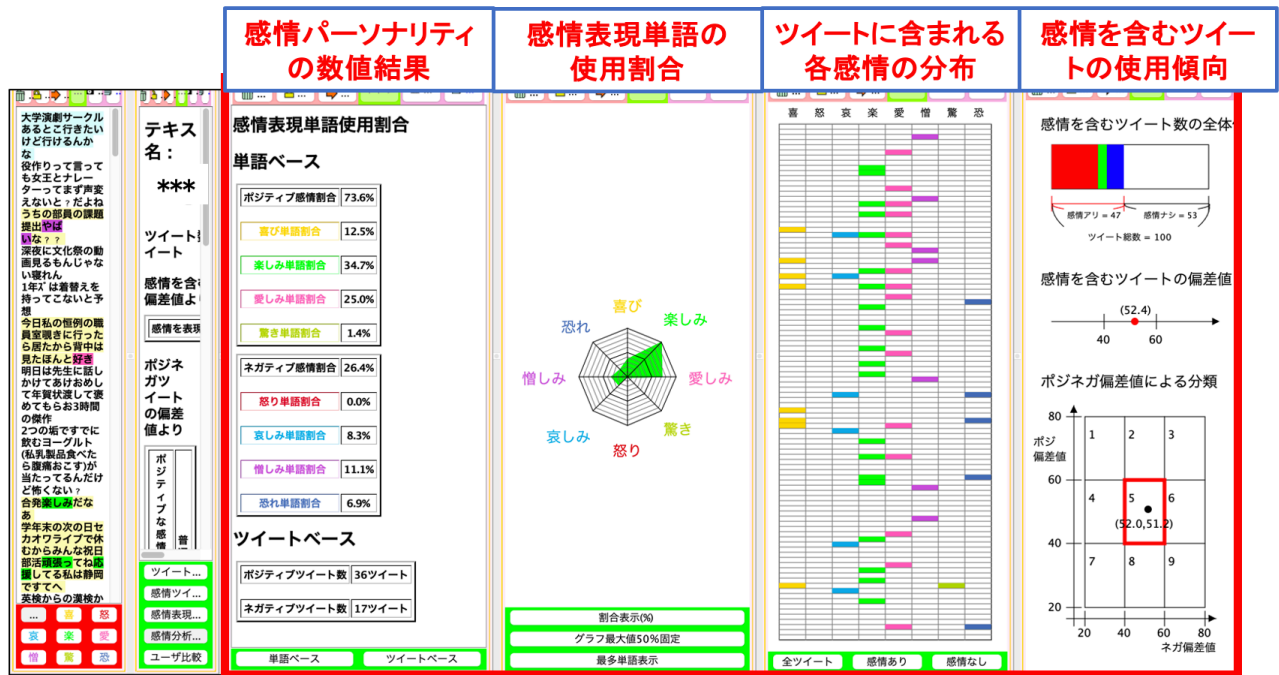


図 4: ユーザ評価支援インタフェース

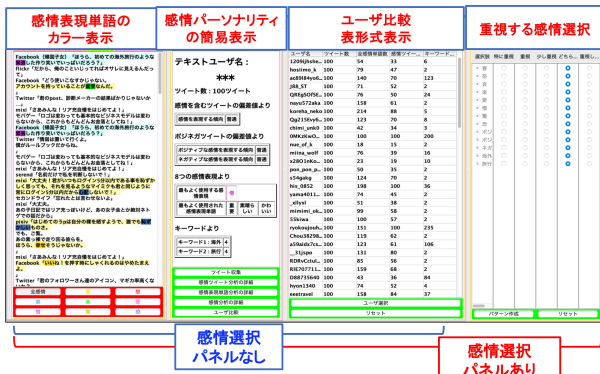


図 5: 実験用ユーザ探索インタフェース

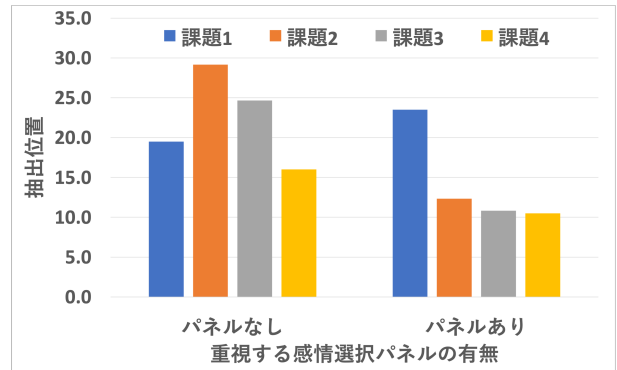


図 6: ユーザの抽出位置

5.2 実験結果

重視する感情選択パネルの有無における表形式表示内でのユーザの抽出位置について図 6 に示す。抽出位置は、表形式表示においてソートを行った後に抽出した 5 名ユーザの位置 (上から 0 番～99 番) の中央値の平均を示している。図 6 において、重視する感情選択パネルを用いることによって、抽出位置を小さくすることが可能になった。この結果からユーザの抽出において、システム利用者の好みを反映させたソート結果を用いることで、抽出したいユーザ候補を表形式表示のより上位に表示することが可能になると考えられる。

課題 1 のパネルありにおいては、あるユーザがソート後に表形式表示の下位のユーザを抽出していた。これについては、ソートを小さい順に表示することも可能であるため、ユーザ探索を支援できると言える。

また、各課題で選択された重みの平均について図 7 に示す。課題 1, 3 において、ポジティブ感情について正の値が得られた。また、課題 2 においては感情表現について重みの平均が小さい傾向があった。さらに、課題 4 においては感情表現に対する重みが負の値を示している結果となった。これらの結果から目的に応じたユーザの抽出において、重視する感情表現が異なることが考えられる。課題 1, 3 ではポジティブ感情を重視

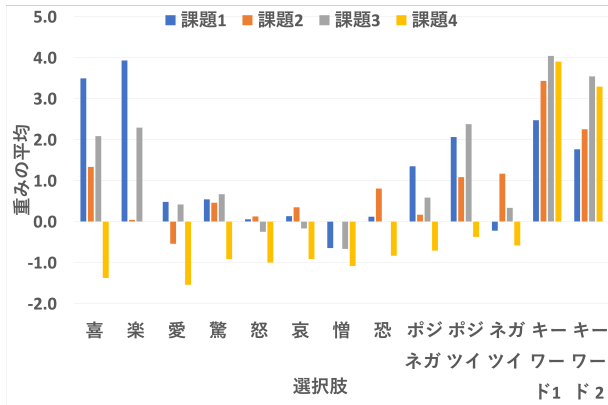


図 7: 各課題で選択された重みの平均

する傾向があるが、探索したいユーザの目的の違いにより課題 1 の方がよりポジティブ感情を重視する傾向が考えられる。また、課題 2 では、平均値は小さい傾向を示しているが、被験者によって重みを正方向または負方向に選択する被験者が分かれる結果が得られたため、平均値は小さくなったと考えられる。さらに、課題 4 に対しては感情表現がユーザの主観を含む情報であるため、客観的な情報を収集したい被験者は感情表現が少ないユーザを探索したい傾向があったと考えられる。

図 8 に抽出した 5 名のユーザ候補を実際にフォローしたいかどうかの判断について、重視する感情選択パネルの有無によるフォローしたいユーザ数平均を示す。ユーザ探索にて抽出した 5 名のユーザについて、重視する感情選択パネルを用いない場合には最低 2 名のユーザをフォローしたい結果を示している。また、フォローしたいユーザ数平均のブレが大きい結果が得られた。一方、重視する感情選択パネルを用いる場合には、安定して 3 名のユーザを評価できる結果を示している。

フォローしたいユーザ数の平均から重視する感情選択パネルの有無によるユーザ探索および評価において、パネルを用いない場合では 2 名のユーザをフォローしたいユーザとして抽出、判断できる。しかし、単純なソートによるユーザの抽出では、課題によってフォローしたいユーザを安定して抽出することができないことが考えられる。一方で、重視する感情選択パネルを用いることによって、安定してフォローしたいユーザを抽出でき、より目的に合うユーザの探索および評価ができると考えられる。

6 おわりに

本研究では Twitter を題材とし、ユーザの感情表現に着目し感情パーソナリティの算出、可視化すること

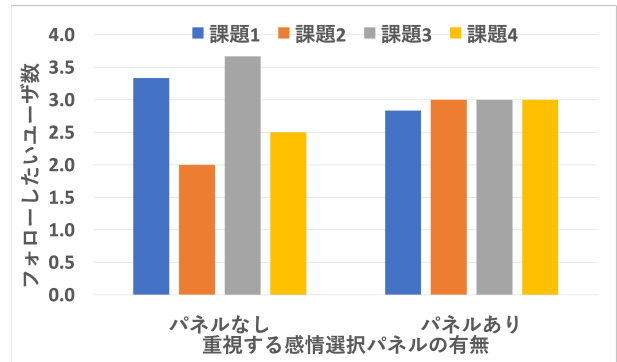


図 8: フォローしたいユーザ数平均

によって、複数のユーザからユーザの探索および評価を支援するためのシステムを構築した。

構築したシステムによって、感情パーソナリティの可視化を行い、ユーザの探索および評価を支援できるかを検証するため検証実験を行った。結果として、被験者の重視する感情を選択せずにユーザの抽出および判断を行った場合、四割のユーザをフォローしたいユーザとして抽出し判断できるが、安定したユーザの抽出を行うことができてなかった。しかし、重視する感情を選択しソートした結果を用いてユーザ抽出および判断したことによって、六割のユーザを目的に合うユーザとして抽出することが可能になった。また、重視する感情の選択結果が探索したいユーザの目的に合わせて異なり、さらにソートによって抽出位置を上位にすることが可能となった。感情パーソナリティの可視化を用いて、ユーザの探索および評価を支援することが可能となった。

今後の課題として、ユーザ探索支援においては、表形式表示におけるソートの固定や抽出したユーザ候補を強調表示し、ほかの複数ユーザと比較し表示することで、ユーザ探索をさらに容易にするためのシステムの改良が求められる。ユーザ評価支援においては、現在 1 ユーザの感情パーソナリティの可視化を行うことで、入力しているユーザの評価を行うシステムになっている。そこで、ユーザ探索で抽出した少数ユーザ (実験時の 5 名のユーザなど) を同時に評価を行えるシステムの改良などが求められる。今後の展望として、感情表現単語の単語数の増加や他の SNS における感情パーソナリティの可視化による支援を行うシステムの発展を検討していきたい。

参考文献

- [1] Twitter, (URL)<https://twitter.com> (2023 年 1 月 27 日確認)

- [2] 水門義之, 田邊洋人, 和泉潔: コロナ禍における Twitter 情報を用いた感情値の計測と人出の関係, 第 36 回人工知能学会全国大会論文集, 1M1-OS-20a-04, pp.1-3(2022)
- [3] Hyo Jin Do, Chae-Gyun Lim, You Jin Kim, Ho-Jin Choi: Analyzing Emotions in Twitter During a Crisis: A Case Study of the 2015 Middle East Respiratory Syndrome Outbreak in Korea, 2016 International Conference on Big Data and Smart Computing (BIGCOMP), No.3, pp. 415-418(2016)
- [4] 北村 太一, 小川 祐樹, 諏訪 博彦, 太田 敏澄: コミュニケーションに着目した Twitter フォロワー ユーザ推薦, 第 26 回人工知能学会全国大会論文集, No.6, pp.1-4(2012)
- [5] 加藤 慎一郎, 濱川 礼: Twitter 感情分析を用いた感情値可視化とユーザ推薦システム, 情報処理学会研究報告, Vol.2013-HCI-152, No.27, pp.1-8 (2012)
- [6] 中村 明: 感情表現辞典, 東京堂出版, (1993).
- [7] 新村 出: 広辞苑第七版, 株式会社岩波書店, (2018)
- [8] 人間の 5 つの基本感情 (46 種類) | ラマ氏とエクマン氏の感情地図, ビジネスのための Web 活用術。個人時代のビジネス力・魅力をあげる戦略学習サイト, (URL)<https://swingroot.com/emotion-type/> (2023 年 1 月 27 日確認)
- [9] 山本 和英: 日本語感情表現辞書, (URL) <http://www.jnlp.org/SNOW/D18> (2023 年 1 月 27 日確認)
- [10] 砂山 渡, 高間 康史, 西原 陽子, 徳永 秀和, 串間 宗夫, 阿部 秀尚, 梶並 知記: テキストデータマイニングのための統合環境 TETDM の開発, 人工知能学会論文誌, Vol.28, No.1, pp.1-12, (2013)
- [11] Twitter Search API, (URL)<https://developer.twitter.com/en> (2023 年 1 月 27 日確認)

StackGAN を用いた感情を表すコミック画像生成に関する 予備的検討

Preliminary Study on Generating Comic Images Expressing Emotion Using StackGAN

¹ 陳 彦嘉 ² 柴田祐樹 ³ 高間康史

Yen-Chia Chen¹, Hiroki Shibata², Yasufumi Takama³

東京都立大学大学院システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University

Abstract: This paper reports the result of preliminary experiments of a method for generating comic images expressing emotions using StackGAN. The proposed method trains StackGAN using the dataset of comics of positive or negative facial images, which are associated with text describing emotions. Generated images are evaluated by applying facial expression recognition. Experiments are conducted changing the types of facial expressions in the training images to investigate the impact of diversity in the training data in terms of emotions on the quality of the generated images.

1 はじめに

本稿では、感情を表現するコミック画像を、StackGAN を用いて生成する手法について予備的検討を行った結果について報告する。企業において、サービスの向上は重要な課題であり、そのヒントとなるのはユーザからのフィードバックとなる。レビューやソーシャルメディアに投稿されたコメントは、企業のサービスに関するユーザの意見が含まれているため、これを活用することでアンケートやインタビューを実施せずにユーザからのフィードバックが得られることが期待できる。

レビューなどのコメントの特徴は、対象サービスなどに対する記述とともに、投稿者の感情が表現されていることである。サービスに満足しているコメントはポジティブな感情を示していることが想定される一方、商品に不満のある場合や、他者を攻撃するような悪意あるコメントの場合はネガティブな感情を示していると考える。従って、レビューコメントを分析する際には、このような感情の観点から分析する必要もあると考える。しかし、コメントを読んで感情の種類やその程度を読み取ることは分析者にとってコストのかかる作業である。また、sentiment analysis [1]などを適用して数値化しても、数値やグラフから感情を認識することは分析者にとって直感的ではないと考える。

本稿では、レビューコメントに含まれる感情を、

表情を表す顔画像として可視化する手法を提案する。特に絵で物語を語るコミックは、登場人物の表情で様々な感情を読者に伝え、感情移入させることが可能であることから、コミック画像として生成することを試みる。提案手法の利用方法として、Charnoff face[2]の様に、多数のレビューコメントから生成したコミック画像を並べて提示することで、その概要把握を支援することが考えられる。また、レビュー投稿前にそれが表す感情をコミック画像として提示することで、その投稿が他者に与える印象を確認させる用途も期待できる。

前述の目的達成に必要な要素技術として、本稿では感情を表すコミック画像の生成手法について検討する。テキストから画像を生成する AI は近年急速に研究が進んでいる分野であり、代表的な生成手法の一つに GAN (Generative adversarial networks, 敵対的生成ネットワーク) [3]がある。GAN は画像を生成する生成器 (generator) と生成画像の真偽を判断する識別器 (discriminator) から構成され、識別器の判断結果をフィードバックして、両者が反復的に学習を行い精度を向上させることで、生成器が高品質の画像を生成可能となる。GAN には様々なバリエーションが存在するが、本稿では感情を表すテキストから画像を生成することを目的とするため、StackGAN[4]を利用する。StackGAN は text-to-image タイプの GAN の中で代表的なものであり、多くの拡張手法が提案されている。画像を生成するタスクを、テキストを

用いて画像を生成するステージ1と、ステージ1の生成結果を用いて最終画像を生成するステージ2の2段階に分けて生成することで高解像度の画像生成を実現しており、細かい線から構成されるコミック画像の生成にも有効であることが期待できる。しかし、StackGANは鳥や花などに関する写真の生成に適用され、その有効性が示されているが、コミック工学においてコミック画像の生成に適用した研究は報告されていない。

提案手法では、ポジティブ、ネガティブな表情のコミック画像に、感情や表情を説明するテキストを付与して学習に用いる。生成した画像の評価において、客観性および再現性を重視するため、感情認識を適用する。学習に用いる画像が表す感情の種類を変更して行った実験結果を比較し、感情に関する学習データの多様性や画像の枚数が生成画像に与える影響を考察する。

2 関連研究

2.1 Stack GAN

テキストから画像を生成する研究のうち、StackGANは代表的な手法であり、拡張手法も提案されている[5][6]。StackGANはステージ1、2の二段階から構成される。ステージ1では、テキストをベクトル化し、画像を生成する。ステージ2ではこの画像を更に入力に用いて、より高解像度の画像を生成して最終的な出力とする。ステージ1は主にテキストの内容を正確に読み取ることを重視して、画像としての質に関してはステージ2で調整する。

StackGANの拡張手法として、StackGANv2[5]は三段階以上のステージにより構成されている。生成に用いる色の一貫性を保つため、正則化項を提案しており、この二点の改良によって、より高解像度の画像の生成が可能となっている。AttnGAN[6]は文単位の入力テキストを一つのベクトルに変換するのではなく、複数の単語に区切り、単語ごとにベクトル化する。これにより、画像は要素ごとに、関連性の高い単語から変換されたベクトルを用いて生成される。

2.2 Russell の感情円環モデル

Russellの感情円環モデルは、ヒューマン・ロボット・インタラクションや感情分析などの工学的研究において良く用いられており、快—不快と覚醒—非覚醒の二軸により構成される平面上に、感情を円環

状に配置したものである。ポジティブな感情には元気、幸せや満足などがあり、快側に配置される。ネガティブな感情は不快側に配置され、緊張や心配、悲しみなどが存在する。平面上に16種類の感情が配置されている。また、これら16感情の外周には、典型的感情エピソードである、驚き、幸せ、怒り、恐怖、嫌気と悲嘆が配置されている。

2.3 GANによるコミック風肖像画生成

顔写真を入力として、コミック風の肖像画を生成する手法が提案されている[7]。この研究では、Pix2Pix[8]とCycleGAN[9]を用いている。入力された顔画像の特徴を残してコミック風の画像に変換するために、複数のタスクに分けて学習を行っている。第一段階ではCycleGANを利用して、コミック画像を用いて顔写真風の画像を生成している。第二段階ではPix2Pixを利用して、生成された画像を元のコミック画像に戻す。最終的に、学習したPix2Pixを利用して、顔写真を入力に用いてコミック画像を生成する。また、作者の異なるコミック画像をデータセットとして用いることで、コミックの画風が学習に与える影響についても考察している。

3 提案手法

3.1 コミック画像データセットの構築

本稿ではManga109¹から、研究用途に提供されているコミックデータセットを利用する。このデータセットには、109タイトルのコミックが含まれており、それぞれ単行本の第一巻が収録されている。また、本研究では使用しないが、登場人物のセリフを、吹き出しの位置と発言者の情報とともに記録したデータも含まれている。

3.2 StackGANによる学習

顔写真とは違い、一枚のコミック画像に含まれている特徴量は少ない。色情報がグレースケールに限定されていること、奥行きに関する情報の欠落により、コミック画像の生成はアニメ風画像の生成よりも困難と考える。また、StackGANをコミック画像の生成に用いた前例も著者らが調査した限り存在しない。そこで本稿では予備的検討として、段階的に画像データセットを増やすことで、StackGANにより生成されるコミック画像がどのように変化するかを検

¹ <http://www.manga109.org/ja/index.html>

証する。

Russell の感情円環モデルの快側の感情から *serene* (晴朗), *contented* (満足), *happy* (幸せ), 不快側から *stressed* (重圧), および典型的感情エピソードから *disgust* (嫌気), *anger* (怒り) を選択し, それらに関する顔画像をコミックデータセットから選び出して学習に用いる. 学習は表1に示すデータセットを用いて3回行う. 表における数値はその感情に対応した画像の枚数である.

表 1: 学習に用いるデータセットの枚数

データ	晴朗	満足	幸せ	重圧	嫌気	怒り
1	2			2		
2	20			20		
3	2	2		2	2	
4	20	20		20	20	
5	2	2	2	2	2	2
6	20	20	20	20	20	20

入力テキストは, 感情の極性, 感情の種類, 表情に関する単語により構成される. 感情の極性を表す単語は, **Positive** (快), **Negative** (不快) のどちらかである. 感情の種類を表す単語は上述の通り, **Positive** の場合は *serene*, *contented*, *happy*, **Negative** の場合は *stressed*, *disgust*, *anger* のいずれかとする. 表情を表す単語は, **Positive** の場合は *laugh* (笑い) と *smile* (微笑み) から, **Negative** の場合は *bellow* (怒鳴る) と *grumpy* (険しい表情) から一単語選ぶ. 感情ごとに, 2種類の表情の画像を同数用意する. 3回の実験において, 学習率 (**learning rate**) は両ステージで 0.0002, 世代数はステージ 1 で 10000, ステージ 2 で 5000 とした.

3.3 生成画像の評価

生成した画像に対して, 生成したい感情の極性を表す表情が生成されているかを評価する. 生成した画像を人手で評価する方法も考えられるが, 本稿では再現性, 客観性を考慮して感情認識を適用して評価する. 感情認識には **Py-Feat**²を利用する. **Py-Feat** が認識する感情のカテゴリは Russell 感情円環モデルとは異なり, 快の感情は *happiness*, 不快の感情は *anger*, *disgust*, *fear*, *sadness*, それ以外の感情は *neutral* と, *surprise* である. **Py-Feat** は各感情に該当する度合いを数値として出力するため, 認識結果で, 度合いが一番高い感情が属する極性を認識結果とする.

4 実験結果

4.1 データセット

画像に関しては, **Manga109** のデータセットから, キャラクターの顔画像部分を 256x256 のサイズで切り出したものを利用する. コミック画像の特性上すべて正面の顔画像にするのは難しく, セリフ (吹き出し) に関しても学習に影響を及ぼす可能性があるため, データセットの拡張も兼ねて左右反転のコピーをする. すなわち, 学習には表1に示した枚数の2倍の画像を用いる.

学習に用いる画像は必ずしもセリフを伴うものではなく, また感情的な単語を含んでいるとも限らない. そこで, 画像に対応するテキスト (3種類の単語) については, 著者の一人が画像から判断して付与する.

4.2 実験結果

4.2.1 実験 1

実験 1 では, 表 1 に示したデータセット 1 を用いて学習する. ステージ 1 の学習では, 世代数 8500 から生成画像がはっきりし, 9000 辺りで崩壊する傾向が見られた. その後のステージ 2 では顔画像が生成されなかったため, 学習に失敗したと考える.

4.2.2 実験 2

実験 2 では, 表 1 に示したデータセット 2 を用いて学習する. ステージ 1 では, 世代数 500 から生成画像がはっきりし, 世代数 7400 と 8800 からそれぞれ約 200 世代崩壊する以外, 安定して顔画像が生成された. ステージ 2 では, 世代数 1000, 2000 で顔の輪郭が生成されたが, 生成する画像にすべてグレイのノイズが発生するため, 感情認識ができなかった.

4.2.3 実験 3

実験 3 では, 表 1 に示したデータセット 3 を用いて学習する. ステージ 1 では, 世代数 750 から生成画像がはっきりし, 世代数 3000 から 3700 の間は崩壊していた. ステージ 2 では, 世代数 250 から生成画像がはっきりした. しかし, 世代数 1100 から 3200 の間は崩壊し, その後 3600 までの 400 世代ではノイズがかかった顔画像が出力された. 以降は再びグレイのノイズ画像しか出力されなかった. 各感情と表情の組み合わせ 8 通りについて, 学習世代数 1000 の生成画像に対する極性判定結果の混同行列を表 2 に

² <https://py-feat.org/pages/intro.html>

示す．行が認識結果に対応する．

表 2：実験 3 の分析結果

感情	晴朗	満足	重圧	嫌気
快				
不快	1		1	
失敗	1	2	1	2

表 2 より，感情が認識できた画像は 8 枚中 2 枚のみであった．認識できた晴朗の画像は，怒り (anger) のスコアが一番高く，快に関する感情は喜び (happiness) のスコアがわずかに出力された．

4.2.3 実験 4

実験 4 では，表 1 に示したデータセット 4 を用いて学習する．ステージ 1 では，世代数 1100 から生成画像がはっきりし，その後生成画像の崩壊は見られなかった．ステージ 2 では，世代数 2400 から，顔の輪郭部分とコマ枠部分にノイズが発生した．はっきりした顔画像は生成されなかったが，ノイズの分布は実験 1，2 と違い，画像すべてにかかるのではなく，外回りに集中していた．

4.2.3 実験 5

実験 5 では，表 1 に示したデータセット 5 を用いて学習する．ステージ 1 では，世代数 1000 から生成画像がはっきりし，5400 から 5800 と 7000 から 7800 の間で崩壊した．ステージ 2 では，世代数 300 から生成画像がはっきりした．2100 から 2800 は生成画像にノイズがかかり，以降はノイズが強くなる傾向が観測された．学習世代数 2000 の生成画像に対する極性判定結果の混同行列を表 3 に示す．

表 3：実験 5 の分析結果

感情	晴朗	満足	幸せ	重圧	嫌気	怒り
快						
不快			1			
驚き						1
中立	2					
失敗		2	1	2	2	1

表 3 より，実験 5 では認識できた 4 枚のうち 3 枚が快の感情を表す画像であった．図 1, 2 に例を示す．Py-Feat ではノイズの有無にかかわらず，輪郭線が細い画像や，デフォルメされたキャラクターに対して正しく認識できない傾向が見られた．

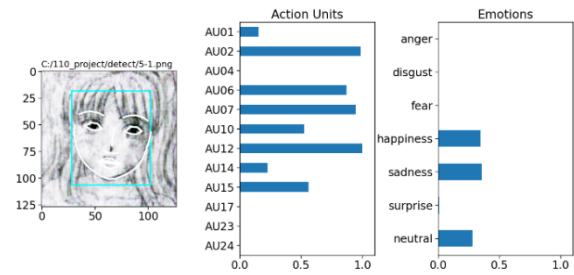


図 1：認識に成功した画像例（実験 5）：
Positive, happy, smile.

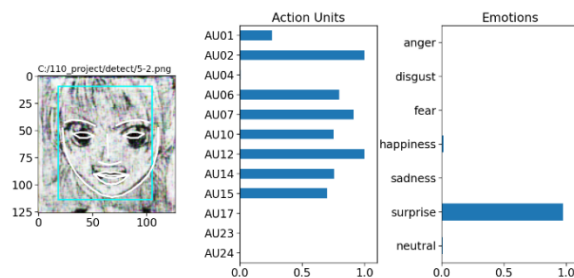


図 2：異なる感情として認識された画像例
（実験 5）：Negative, anger, bellow.

実験 3 と 5 において，同じセリフから生成された同一キャラクターの画像を図 3, 4 に示す．表 2, 3 に示した通り，実験 3 では不快の感情が，実験 5 では快の感情が生成しやすい傾向が見られたが，今後生成枚数を増やして確認する必要があると考える．

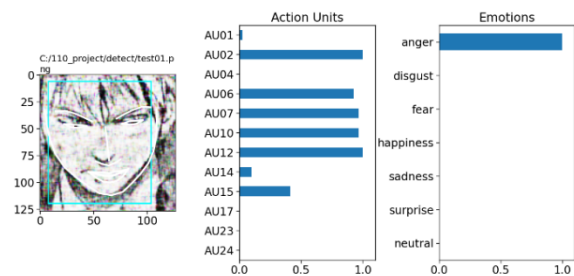


図 3：生成画像例（実験 3）：
Negative, stressed, grumpy.

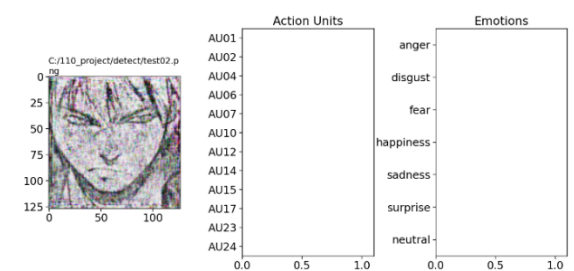


図 4：生成画像例（実験 5）：
Negative, stressed, grumpy.

実験 3 と 5 で生成された、感情を認識できた画像のうち、同じキャラクターの画像が一種類存在した(図 5, 6)。これより、画像として同様の表情に見えるが、感情認識の結果は大きく異なっており、感情認識の頑健性は低いといえる。

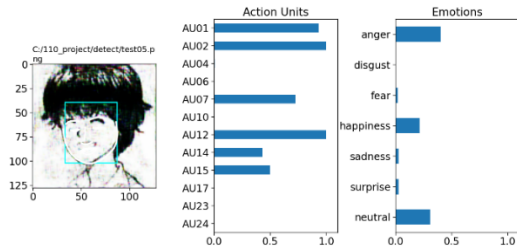


図 5 : 生成画像例 (実験 3) :
Positive, serene, laugh.

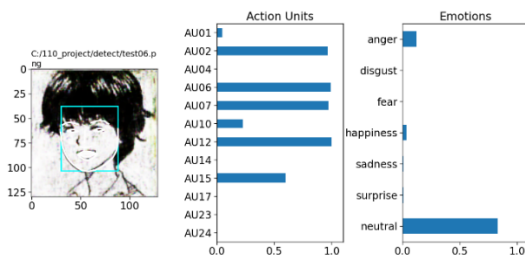


図 6 : 生成画像例 (実験 5) :
Positive, serene, laugh.

4.2.3 実験 6

実験 6 では、表 1 に示したデータセット 6 を用いて学習する。ステージ 1 では、世代数 200 から生成画像がはっきりし、4700 から 5200, 5700 から 8300 の間と、9400 以降で崩壊した。ステージ 2 では、世代数 3000 から、実験 4 と同様のノイズ画像が生成され、表情認識が適用できなかった。

4.2.4. 学習データの多様性による影響の考察

感情の種類や、学習に用いる画像の枚数と認識結果の関係について考察する。感情の種類と生成画像の関係については、2 種類、4 種類と 6 種類の感情に対して、それぞれ 2 回の実験を行った結果、2 種類の感情を用いた学習では、2 回の実験どちらも画像が上手く生成されなかった。これより、感情の種類はある程度用意する必要があると考える。

学習に用いる画像の枚数と認識結果の関係については、感情認識ができた実験 3, 5 は各感情 2 枚ずつ学習に用いた場合であり、20 枚ずつ用いた場合は一度も感情認識が成功しなかった。この結果より、画

像を増やしても生成画像の質は向上しない可能性が示されたが、今後さらに枚数を増やしての検証が必要と考える。また、学習過程は単調ではなく、一度表情が読み取れる画像が生成されるようになった後、世代を重ねると崩壊する場合が観測された。この現象についても、今後さらに検証が必要と考える。

Py-Feat の認識結果については、快の感情に分類された画像はなく、不快の感情が 3 枚、中立 (neutral) が 2 枚、驚きが 1 枚という結果になった。図 5 と図 6 では顔の領域として検出された位置がずれている。これは、図 5 で眉毛の部分が上手く生成できてなく、目の部分が眉毛に誤認されたためと考える。また、図 1, 2, 3, 4 を見る限り、多少のノイズがあっても表情が認識できる場合があるため、キャラクターの画風、輪郭線、描かれている角度の影響が認識精度に与える影響が大きいと考える

5 まとめ

本稿では、感情を表現するコミック画像を、StackGAN を用いて生成する手法について予備的検討を行った。実験を行った結果、4 種類以上の感情に対してそれぞれ 2 枚の画像を与えた場合に感情認識が可能であったことを示した。また、学習過程は単調ではなく、後半に生成画像が崩壊する現象が多くみられたことから、安定した画像生成にはまだ至っていないと考える。Py-Feat による感情認識結果は、快の感情を認識できない傾向にあるが、顔として検出された領域を見る限り、キャラクターの顔輪郭と顔パーツの位置はある程度正確に把握可能であることを確認した。

本稿で示した実験結果より、ステージ 2 での学習が失敗しやすい傾向がみられたため、安定した画像生成のために今後は、学習率の調整、及びモデル構造の変更を検討する予定である。また、テキスト入力の多様性向上や、ディフュージョンモデルを用いた同様の実験に取り組む必要があると考える。

参考文献

- [1] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal*, Vol. 5, No. 4, pp. 1093-1113 (2014).
- [2] Christopher J. Morris, David S. Ebert, and Penny L. Rheingans. "Experimental analysis of the effectiveness of features in Chernoff faces." 28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making (2000).

- [3] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. “Generative Adversarial Networks: An Overview.” IEEE Signal Processing Magazine, Vol. 35, No. 1, pp. 53-65 (2018).
- [4] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N. Metaxas. “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.” ICCV2017, pp. 5907-5915 (2017).
- [5] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N. Metaxas. “StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks.” IEEE Trans, on Pattern Analysis and Machine Intelligence, Vol. 41, No. 8, pp. 1947-1962 (2018).
- [6] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. “AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks.” CVPR2018, pp. 1316-1324 (2018).
- [7] Yen-Chia Chen, Lieu-Hen Chen, Hiroki Shibata, and Yasufumi Takama. “Styled Comic Portrait Synthesis Based on GAN.” JSAI2021: Advances in Artificial Intelligence, Springer, pp. 69-80 (2022).
- [8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. “Image-to-Image Translation with Conditional Adversarial Networks.” CVPR2017, pp. 1125-1134 (2017).
- [9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.” ICCV2017, pp. 2223-2232 (2017).

GANによる質問答弁の対話関係評価を導入した 議会会議録の要約文章生成

Summarization of parliamentary proceedings introducing evaluation of question-answer dialogue relationships by GAN

大野木裕睦^{1*} 森辰則¹
Hiromu Onogi¹ Tatsunori Mori¹

¹ 横浜国立大学大学院環境情報学府

¹ Graduate School of Environment and Information Sciences, Yokohama National University

Abstract: There is a great demand for automatic summarization of parliamentary proceedings. It is necessary considering the correspondence between questions and answers when summarizing, but an effective method has not yet been established. We tested a new method to train a T5-based generative summary model to consider the correspondence between questions and answers by using an adversarial generative network (GAN). We also introduced a new framework for the loss function of the generator and tried a novel method that uses the generated question summaries as input when generating answer summaries. The summaries generated by the proposed method obtained higher accuracy than the ones of many existing models. Furthermore, we believe that the "readability" of the generated summaries was improved as a result of human evaluation.

1 はじめに

政治課題に関する議論は、議会において行われており、会議録には議論の質問や答弁が記録されている。しかし会議録は、議会での発言を書き起こされただけでまとめられておらず、読みづらいという問題がある。そのため「議会だより」等の文書に、会議録の要約が掲載されるが、50字程度と短く、人手で作成するには時間とコストが発生する。そこで、精度の高い議会会議録の自動要約システムが求められている。

議会会議録の自動要約には、次のような課題がある。(i) 先行研究は従来、抽出型要約手法を採用してきたが、要約ペアを作る上で文抽出型の手法には限界がある上、昨今は生成型要約手法の性能向上が著しいため、これを試すべきである点と、(ii) 質問答弁の対話関係を考慮しながら要約する必要があるという点である。

(i) の点に対して、我々は T5 を用いて生成型要約モデルを構成・訓練する。(ii) の点に対して、我々は質問と答弁の要約ペアを生成するモデルと、その対話関係の妥当性を検証するモデルを同時に訓練していくことが有効ではないかという仮説を立てる。このような枠組みを、敵対的生成ネットワーク (GAN) を利用して

実現する。生成器 G は会議録の質問発言と答弁発言を入力して質問と答弁の要約ペアを生成する。一方で識別器 D は、質問要約と答弁要約のペアを受け取り、それが人手で作られたものか、あるいは、生成器 G により生成されたものか、を判別する。これら両モデルを交互に学習させることで、生成器 G は識別器 D を騙せるほど精度の高い要約生成をするよう洗練されていき、結果的に自然な対話構造のある質問要約・答弁要約のペアを出力することを期待する。また、生成器 G の損失関数について、教師強制を行いつつ、識別器 D の損失を加味するような新たな枠組みを導入する。

さらなる性能向上を目指して、次のような新規手法を試す。(1) 対話構造を加味しながら要約生成を行うために、生成された質問要約を、答弁要約生成の入力として利用することで、対話の流れを考慮させる。この枠組みをカスケード型生成器と呼ぶ。(2) 質問要約生成・答弁要約生成それぞれについて、入力に質問・答弁両方の原文を与える工夫を行う。(3) 原文と要約を入力し、与えられた要約が人手で書かれた要約か、生成器 G により生成された要約かを判別する識別器 D_{QPS} 、 D_{APS} を導入する。これにより、原文と要約の対応を考慮できるようになることが狙いである。

東京都議会会議録に対して、提案手法を利用し実験を行う。会議録を訓練用データと評価用データに分割し、モデルの訓練と評価を実施する。評価にあたって

*連絡先： 横浜国立大学大学院環境情報学府
〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-1
E-mail: ridivarg@outlook.com

は、自動評価および人手評価を行う。

2 関連研究

議会会議録を自動要約する試みは既に行われている。例えば NTCIR QALab-PoliInfo という評価型ワークショップにおいて、地方議会会議録の要約を行うタスクが設定されている。これに関連して、多くの先行研究がなされている [4][5][2]。

2.1 要約手法の種類

自動要約手法は、抽出型要約と生成型要約に大別される。抽出型要約は要約対象の文章から、代表文を抽出する手法であるのに対し、生成型要約は文の表現を抽象化し、1 から要約を生成する手法である。

議会会議録の要約においては、長らく抽出型要約手法が主流であった [4][5]。しかし、抽出型要約は、質問と答弁の対話関係を保ちつつ、質問答弁それぞれの代表文を抽出することが難しい。

一方で生成型要約は、文法が正確で、文脈に整合性のある要約を生成することが難しいとされていた。しかし近年、生成型要約が目覚ましい発展を遂げている。先述した課題をほとんど克服しつつあり、人間の性能にどんどん迫っている。代表的な生成型要約モデルのほとんどは、とりわけ Transformer ベースの事前学習モデルをファインチューニングすることで構成される。例えば、GPT や BART、T5[6] などが挙げられる。

2.2 対話構造を把握することの重要性

木村ら [7] は、議会会議録の要約の実現において、精度向上のために質問と答弁の対応関係を把握することが重要と指摘している。しかし、先行研究の多く ([5] 等) は質問と答弁の対話構造を十分に捉えられておらず、いまだ質問答弁の対話構造を把握するための確立された手法は無い。

2.3 GAN を用いた文書要約

本論文では、敵対的生成ネットワーク (GAN)[1] を用いた議会会議録の要約モデルを提案する。自動要約において GAN を利用する手法は、Liu ら [3] により提案されているが、この手法は単テキストの要約のために GAN を導入することで、品質向上を図るものである。一方で我々は、2つのテキスト間の対応関係を勘案するため、GAN を導入する点に違いがある。

3 GAN による対話関係評価を導入した生成型要約モデル

3.1 設計方針

2 章で、(i) 生成型要約モデルを試す必要があると考える点 (2.1 節)、(ii) 文脈や論理展開を考慮しながら要約する必要があるという点 (2.2 節) を指摘した。これを踏まえ、以下のような方針を立てた。

3.1.1 生成型要約モデルの利用

i. に関して、我々は、T5 ベースの日本語事前学習モデルをファインチューニングし、生成型要約モデルを構成する。T5 モデルを選択する理由は、通常の Transformer とは異なり Relative-Positional-Encoding を利用しているため、実質的に入力長に制限が存在しないからである。長大な会議録を扱う際に、入力長に制限があるに対応できないケースがあり得る。

3.1.2 GAN の枠組みを用いたモデルの構築

ii. に関して、我々は質問と答弁の要約ペアを生成するモデルと、その対話関係の妥当性を検証するモデルを同時に訓練していくことが有効ではないかと仮説を立てる。これによって、対話関係の「自然さ」を担保する狙いである。さらにこの際、両モデルを敵対的生成ネットワーク (GAN) の枠組みにおいて共同で訓練することを考える。

3.2 基本設計

敵対的生成ネットワーク (GAN) を利用し生成型の要約生成を行う手法の基本設計について述べる。基本設計には、生成器 G_Q 、 G_A の訓練において、人手要約で教師強制しながらも、識別器 D の損失を利用して学習を進める手法 (3.2.5 節) のような新しい枠組みを含んでいる。また本研究においては、基本設計に加えて、さらなる新規手法も試す。これらについて、3.3 節、3.4 節、3.5 節にて詳述する。

3.2.1 GAN による訓練

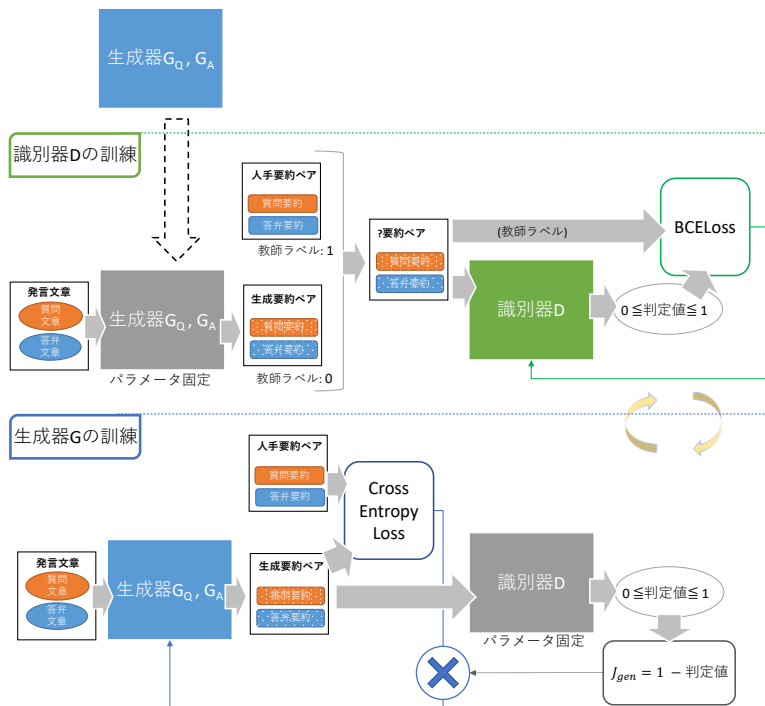
GAN による訓練の模式図を図 1 に示す。

まず、会議録コーパスから質問原文・答弁原文を用意し、さらにそれらについての人手による要約 (人手要約) も用意する。

生成器は質問要約用生成器 G_Q 、答弁要約用生成器 G_A に分かれている (3.2.3 節に詳述)。それぞれ会議録

GANによる訓練

通常の教師強制でファインチューニング済みの生成器



実際の運用



生成器 G_Q, G_A のみ利用する。識別器 D は利用しない。

図 1: GAN による訓練と、生成器の実運用についての模式図

の質問原文・答弁原文を入力として受け付け、それをもとに生成質問要約・生成答弁要約を出力するモデルである。GAN における訓練については、生成要約と人手要約のクロスエントロピー損失を算出し、この損失に識別器の出力を掛け合わせ、フィードバックすることで進める (3.2.5 節)。ただし、GAN 訓練前に、人手要約を用いた教師あり学習でファインチューニングを行っておき、これをもとに GAN で学習する (3.2.6 節)。

一方で識別器 D は、質問要約と答弁要約のペアを受け取り、それが人手で作られたものか、あるいは、生成器 G_Q, G_A が自動生成したものか、を判定する (3.2.4 節)。

これらを交互に学習させることで、生成器は識別器を騙せるほど精度の高い要約生成をするようになるよう洗練されていき、結果的に自然な対話構造のある質問要約・答弁要約のペアを出力することを期待する。つ

まり、手本となる人手の要約により近い要約を生成することを期待している。

3.2.2 GAN によって得られるものはなにか

識別器 D は、ひとまず人間生成の要約か、機械生成の要約かという判断だけで、内容の良さは評価していない。一般に機械翻訳や生成型の自動要約では、流暢さ、重要内容の保持、その他で生成されたテキストの良さを計測するが、上記の識別器の設計だと、原文を与えない方式なので、「流暢さ」がおもに判断されるのではないかと推測する。

3.2.3 生成器 G_Q, G_A の詳細

生成器 G は、質問要約用生成器 G_Q 、答弁要約用生成器 G_A に分かれている。生成器 G_Q, G_A は、会議録の原文を入力として受け付け、それをもとに生成質問要約・生成答弁要約を出力する。

質問要約用生成器 G_Q 、答弁要約用生成器 G_A ともに、Huggingface Model Hub 上に公開されている T5 事前学習モデル¹を利用して構成する。

運用上、生成器 G の入出力テキストは、すべて日本語 T5 用トークナイザー (SentencePiece) によるトークン列である。

GAN における生成器 G の訓練は、識別器 D による判定値を損失としてフィードバックすることで行う。ここで我々は、生成器 G を人手要約によって教師強制しながら、識別器 D から得られた損失をフィードバックして学習を進める新手法を導入する。3.2.5 節にて詳述する。

3.2.4 識別器 D の詳細

識別器 D は「生成質問要約と生成答弁要約のペア」(生成要約ペア) あるいは、「人手質問要約と人手答弁要約のペア」(人手要約ペア) を受け付ける。出力は、与えられた要約ペアが人手により書かれたものか、生成器 G_Q, G_A により生成されたか、の判定値である。

$$D(\text{質問要約, 答弁要約}) = \text{判定値} \quad (0.0 \leq \text{判定値} \leq 1.0)$$

判定値については、0 に近いほど生成要約ペア、1 に近いほど人手要約ペアらしいとする。

識別器 D は、東北大学が Huggingface Model Hub 上で公開している日本語 BERT 事前学習モデル²を利

¹<https://huggingface.co/sonoisa/t5-base-japanese>

²<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

用し、これを転移学習 (ファインチューニング) することで構成する。

運用上、識別器 D の入力テキストは、日本語 BERT 用トークナイザー (MeCab) によるトークン列である。

日本語 BERT 事前学習モデルを、生成要約ペアと人手要約ペアを与えることでファインチューニングし、識別器を訓練する。ファインチューニングは、人手要約の教師ラベルを 1、生成要約ペアの教師ラベルを 0 とする教師あり学習で行う。特別トークン [SEP] で質問要約と答弁要約を分割して入力する。

3.2.5 GAN の訓練における生成器 G_Q 、 G_A の損失関数

従来の GAN に基づく設計では、生成器 G_Q 、 G_A の訓練を次のような仕組みで行う。

$$\begin{aligned} \text{判定値} &= D(G_Q(\text{質問原文}), G_A(\text{答弁原文})) \\ BCELoss(1, \text{判定値}) &=: \text{生成器 } G_Q, G_A \text{ の損失} \end{aligned}$$

しかし、本研究で設計する生成器 G_Q 、 G_A と識別器 D の入出力の形式は大きく異なっている。したがって、生成器 G_Q 、 G_A と識別器 D を連続したニューラルネットワークとして扱っても、損失を逆伝播させる際に障害が生じる。ゆえに、生成器 G_Q 、 G_A と識別器 D を独立したモデルとして扱いつつ、生成器 G_Q 、 G_A に識別器 D の損失をフィードバックさせる方法が必要である。

我々はこの課題に対して、次のような新規手法を適用することで対処する。生成器 G_Q 、 G_A の訓練において、人手要約で教師強制しながらも、識別器 D の判定値を利用して学習を進める。

$$\begin{aligned} J_{gen} &= 1 - D(G_Q(\text{質問原文}), G_A(\text{答弁原文})) \\ CrossEntropyLoss(G_Q(\text{質問原文}), \text{人手質問要約}) \\ &\times J_{gen} \quad =: \text{生成器 } G_Q \text{ の損失} \\ CrossEntropyLoss(G_A(\text{答弁原文}), \text{人手答弁要約}) \\ &\times J_{gen} \quad =: \text{生成器 } G_A \text{ の損失} \end{aligned}$$

通常、T5 モデルの訓練時にはクロスエントロピー損失を利用するが、本手法ではクロスエントロピー損失に J_{gen} を掛け、これを生成器の損失とする。ここで J_{gen} は、人手要約ペアのラベル (=1) と生成要約ペアの判定値の差分、すなわち「本物の要約ペア」に対する「偽物の要約ペア」の損失と見なすことができる。

この手法では、もとのクロスエントロピー損失に対してスカラー値である J_{gen} を乗算しているだけなので、損失の勾配の大きさは制御されるが、勾配の向きは制御されない。しかし、損失が多層 NN を逆伝播するなかで、各ニューロンにおける勾配ベクトルは複雑に変

化していくはずであるため、有効な手法であると期待する。

画像生成 GAN 等においては、ランダムノイズを生成器に入力して訓練を行うため、このような訓練手法を導入できない。一方で、文章要約生成は、明確な教師データが存在しているため、この訓練手法を導入できる。

生成器の損失関数を以上のように定義する手法は、我々の知る限り無い。

3.2.6 GAN 訓練前に生成器 G_Q 、 G_A をファインチューニングする手法

生成器 G_Q 、 G_A の学習が不十分な場合、識別器 D が質疑応答の対応関係を見逃して生成文章の自然さのみで判断する可能性があることが懸念される。この対処として、先に生成器 G_Q 、 G_A のみ、人手要約を用いた教師あり学習でファインチューニングを行う。これをもとに、GAN で学習する。

3.3 カスケード型生成器

上述の生成器の設計では、質問要約と答弁要約を独立に生成しているため、質問と答弁の対話関係を加味しながら要約しているとは言い難い。これを加味するには、独立せずに 2 テキストを出力できる、すなわち質問要約と答弁要約のペアを出力できるようなモデルが必要である。しかし、この条件を満たすような Transformer ベースの事前学習モデルは、今のところ存在しない。

そこで、質問要約生成器 G_Q が生成した質問要約を、答弁要約生成器 G_A に入力することにより、「質問要約を加味して答弁要約を生成できる」ため、対話の流れに則って、質問答弁要約ペアを生成できると考えた。以下、この仕組みを導入した生成器を「カスケード型生成器」と呼ぶ。また、略称 **cascade** でも示す。

$$G_Q(\text{質問原文}) = \text{生成質問要約}$$

$$G_A(\text{答弁原文}, \text{生成質問要約}) = \text{生成答弁要約}$$

カスケード生成器の枠組みについて、図 2 に示す。

より詳細には、T5 に特別トークン <q_sum> を導入して、表 1、表 2 のように入出力を行う。

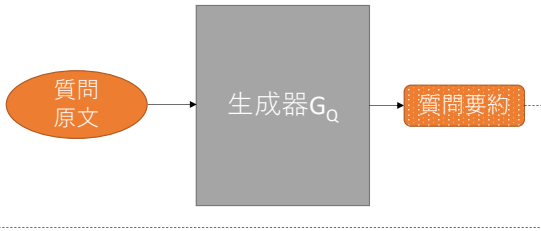
表 1: 質問要約を生成 (Step1)

入力	出力 (目標)
質問原文	生成質問要約

表 2: 生成質問要約を利用し、答弁要約を生成 (Step2)

入力	出力 (目標)
答弁原文 <q_sum> 生成質問要約	生成答弁要約

Step1 質問要約生成



Step2 生成質問要約を再帰的に利用し、答弁要約生成

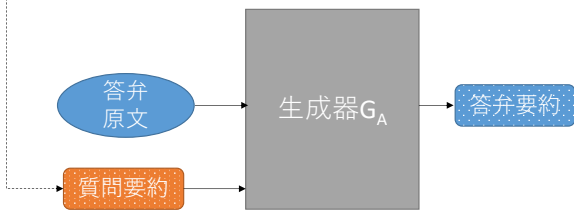


図 2: カスケード型生成器

3.4 各生成器に質問原文と答弁原文の両者を入力する手法

質問要約生成・答弁要約生成それぞれについて、入力に質問・答弁両方の原文を与えるほうが良い可能性を考え、検証する。

$$G_Q(\text{質問原文}, \text{答弁原文}) = \text{生成質問要約}$$

$$G_A(\text{質問原文}, \text{答弁原文}) = \text{生成答弁要約}$$

より詳細には、T5 に特別トークン<question>、<answer>を導入して、表 3、表 4 のように入出力を行う。以下、この手法について、略称 **SET** で示す。

表 3: G_Q に、質問・答弁原文両方を入力する場合

入力	出力
<question>質問原文<answer>答弁原文	質問要約

表 4: G_A に、質問・答弁原文両方を入力する場合

入力	出力
<question>質問原文<answer>答弁原文	答弁要約

3.5 原文と要約の対応を判定する識別器 D_{QPS} 、 D_{APS} の導入

上述の識別器 (3.2.4 節) は質問要約と答弁要約のペアを受け付けるが、原文を考慮しない。したがって、原文と要約の対応を加味できない。追加の識別器として、質問と答弁それぞれについて、原文と要約を入力し、与えられた要約が人手要約か生成要約か判定するようなモデルを構成する。

$$D_{QPS}(\text{質問要約}, \text{質問原文}) = \text{判定値} (0.0 \leq \text{判定値} \leq 1.0)$$

$$D_{APS}(\text{答弁要約}, \text{答弁原文}) = \text{判定値} (0.0 \leq \text{判定値} \leq 1.0)$$

判定値については、0 に近いほど生成要約、1 に近いほど人手要約らしいとする。

以下、質問原文と質問要約の対応を判定する識別器を D_{QPS} 、質問原文と質問要約の対応を判定する識別器を D_{APS} と呼称する。 D_{QPS} 、 D_{APS} と、先述した識別器 D を併用し、検証を行っていく。

なお、これらを併用する際、生成器 G_Q 、 G_A の訓練は次のようになる。

$$J_1 = 1 - D(G_Q(\text{質問原文}), G_A(\text{答弁原文}))$$

$$J_2 = 1 - D_{QPS}(G_Q(\text{質問原文}), \text{質問原文})$$

$$J_3 = 1 - D_{APS}(G_A(\text{答弁原文}), \text{答弁原文})$$

$$\text{CrossEntropyLoss}(G_Q(\text{質問原文}), \text{人手質問要約})$$

$$\times J_1 J_2 J_3 \quad \text{:= 生成器 } G_Q \text{ の損失}$$

$$\text{CrossEntropyLoss}(G_A(\text{答弁原文}), \text{人手答弁要約})$$

$$\times J_1 J_2 J_3 \quad \text{:= 生成器 } G_A \text{ の損失}$$

以下、この手法について、略称 **AQPS** で示す。

4 評価実験

実際の地方議会会議録 (平成 23 年 6 月から平成 27 年 3 月までの東京都議会会議録) に対して、提案モデルがどれほど有効なのか評価する。NTCIR-15 QALab-PoliInfo2³ Dialog summarization task の設定するデータセット、及び評価手法を利用して実験を行う。

4.1 データセット

東京都議会における一般質問および代表質問の概要は都議会だよりに掲載されている。都議会だよりは、議会で記載された内容が議会事務局の職員により作られていることから、人手により作成された「正解の要約」とみなすことができる。また、都議会だよりでは、質問項目ごとに質問者と答弁者が示されており、質問とそれに対応する答弁が簡潔にまとめられている。そこで、都議会だよりに記載された質疑の要約 (50 字程度) を正解として用いる。

データは全 692 個の出題項目からなり、それぞれには質問原文と答弁原文のペア、人手による質問要約と答弁要約のペアが含まれている。438 個を訓練用データとして扱い、残り 254 個を評価用データとして扱う。

³<https://poliinfo2.github.io/>

4.2 マイナーチェンジモデルの訓練

表5のように追加の新規手法を適用したマイナーチェンジモデルも訓練する。

表 5: マイナーチェンジモデル

cascade	×	×	×	×	○	○	○	○
SET	×	×	○	○	×	×	○	○
AQPS	×	○	×	○	×	○	×	○

4.3 学習のハイパーパラメータ

生成器 G_Q, G_A 訓練時のバッチサイズは1、エポック数は3である。

識別器 D, D_{QPS}, D_{APS} 訓練時のバッチサイズは6、エポック数は4である。

GANの訓練においては、生成器の学習と識別器の学習を、交互に25回繰り返す。

4.4 評価

評価用データを対象に要約を生成し、これと人手要約の差を ROUGE スコアによって自動評価し、比較する。また11個の評価用データを無作為に取り出し、これに対応する生成要約を10人に人手評価してもらった。(6人: 情報工学を専攻する学生、1人: 気象関係の研究者、1人: デザイナー、1人: 動画配信者、1人: 主婦) 評価観点は表6のとおりである。

表 6: 人手評価の評価観点

Q-content	質問部分について、人手要約と内容の合致
A-content	答弁部分について、人手要約と内容の合致
Well-formed	文法的な正しさ
Dialog goodness	質問と答弁の対話関係が成立しているか
Overall goodness	質問・答弁要約ペアとしての全体的な良さ

人手評価の対象は、次の5つのモデルに限定した。

- モデル1(ベースライン): 入力を原文、出力を人手要約とするよう、通常の教師強制でファインチューニングした T5 モデル (質問要約生成器、答弁要約生成器)

表 7: 人手評価の結果

	Q-content	A-content	Well-formed	Dialog goodness	Overall goodness	(Q-Rouge)	(A-Rouge)
モデル1(ベースライン)	1.308	1.503	2.562	1.921	1.528	0.275	0.247
モデル2(_cascade.AQPS)	1.892	1.595	2.708	2.025	1.707	0.300	0.301
モデル3(_SET.cascade)	1.508	1.558	2.437	1.756	1.453	0.296	0.291
モデル4(_SET.AQPS)	1.608	1.595	2.321	1.974	1.657	0.304	0.294
モデル5(_SET.cascade.AQPS)	1.808	1.641	2.491	2.053	1.798	0.299	0.290

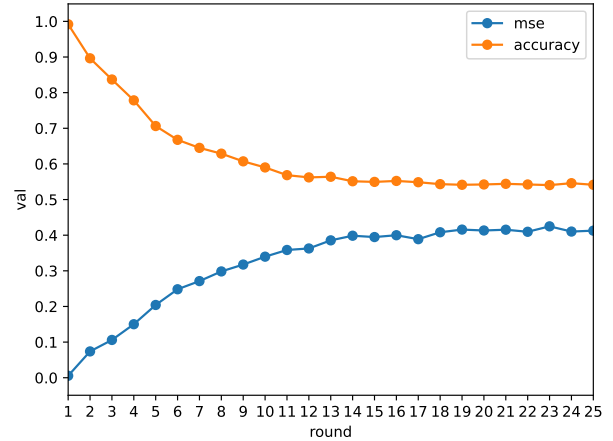


図 3: 識別器 D の性能推移

- モデル2~5: マイナーチェンジモデルのうち、4つ(表8)

表 8: 人手評価対象のマイナーチェンジモデル

	モデル2	モデル3	モデル4	モデル5
cascade	○	○	×	○
SET	×	○	○	○
AQPS	○	×	○	○

評価結果を、表7に示す。

5 考察

5.1 GANの訓練における、生成器 G_Q, G_A への新たな枠組みの損失関数の導入

識別器 D が GAN の学習を繰り返す中で、どのように性能が推移するかを図5に示す。段々と正解率 (accuracy) が0.5に向かっており、識別器が生成要約と人手要約を区別できなくなっていることがわかる。また、平均二乗誤差 (mse) も、段々と0から大きくなっている。したがって、生成器 G_Q, G_A によって識別器 D が騙されるよう、学習が進んでいることがわかる。

ゆえに、3.2.5 節で述べたような、生成器 G_Q 、 G_A の訓練において、人手要約で教師強制しながらも識別器 D の損失を利用して学習を進める新規手法を導入しても、生成器と識別器とで敵対的に学習が進むと言える。

5.2 先行研究との比較

評価型ワークショップ NTCIR-15 QALab-PoliInfo2 Dialog summarization task に提出されたモデルと、我々が訓練したモデルの ROUGE スコアを表 9 に示す。なお表中の、JRIRD、nukl、TO、Forst は、本 task へ他の参加者が提出したモデルを表す。

表 9: 各モデルの ROUGE スコア

モデル	ROUGE
JRIRD	0.321
モデル 2(.cascade_AQPS)	0.301
基本設計	0.296
モデル 5(.SET_cascade_AQPS)	0.295
モデル 1(ベースライン)	0.261
nukl	0.258
TO	0.244
Forst	0.241
(本 task 参加者の平均)	0.185

我々の制作したモデルは、他のモデルの平均を大きく超えた。基本設計であっても、高い水準を獲得しており、GAN の導入が、要約生成の品質改善に寄与していることがわかる。

表 10: モデルの出力例

	質問	答弁
原文	我が国の経済は、自由民主党政権による景気対策や金融対策により、少しずつ明るい兆しが見え始めています。こうした景況の改善が、東京の産業を支えている数多くの中小零細企業にしっかりと行き渡る施策の展開こそが都政の重要な役割であると考えます。また、東京の経済の中長期的な発展を確実なものとするためには、中小企業の成長を行政として戦略的にサポートすることも不可欠です。現政権のアベノミクスの打ち出す成長戦略と連携して、東京の産業振興を着実に推し進めて、中小零細企業の活力を高め、雇用就業の場も生み出していくという総合的な政策の展開こそが必要となっています。都は、中小零細企業への支援を初め、東京の産業振興に向けてどう取り組んでいく考えか、知事の所見を伺います。	次に、雇用就業対策についてであります。東京の産業の成長と都民生活の安定を図るため、少子高齢化の影響を踏まえ、労働力を確保し、一人一人の能力や個性を生かした働き方を実現することが重要であります。現在、非正規雇用で働かざるを得ない若者や、出産等で仕事を離れた女性、これまでの経験等を生かして働くことを目指す高齢者に対して、就業の推進に向け、きめ細かく支援することが課題となっております。このため都は、正社員を希望する若者への重点的な就業支援の実施や、女性の再就職を後押しする新たな取り組みの展開、高齢者が希望や能力に応じて働くための支援の強化など、雇用就業施策の一層の充実を図ってまいります。今後とも、産業の成長を支え、多様な人材が活躍できる雇用就業の仕組みづくりを着実に進めてまいります。
人手要約	若者や女性、高齢者が能力を十分発揮し働ける社会の仕組みづくりを。	きめ細かい支援が課題。能力や個性を生かした働き方の実現へ施策を充実。
モデル 1(ベースライン)	活力を高め雇用就業の場も生み出していく政策の展開が必要。所見は。	雇用就業の仕組みづくりを着実に進め、多様な人材が活躍できる雇用就業の仕組みづくり。
モデル 3(.SET_cascade)	中小零細企業への支援をはじめ東京の産業振興にどう取り組むか。	若者の重点的就业支援実施や高齢者の再就職支援等、雇用就業施策の一層の充実を図る。
モデル 5(.SET_cascading_AQPS)	[1] 東京の産業振興に向けどう取り組むのか。[2] 中小企業の活力を高め雇用就業の場も生み出していく総合的な政策展開が必要。所見は。	[2] 若者の重点的な就業支援の実施や高齢者の希望や能力に応じて働くための支援の強化等雇用就業施策の一層の充実を図る。

5.3 質問文章と答弁文章をまとめて入力する手法 (SET)

モデル 2(.cascade_AQPS) とモデル 5(.SET_cascade_AQPS) を比較し、SET の有効性を検討する (表 7)。Q-Content 指標、すなわち質問要約内容の精度は減少してしまった。このことから、質問要約生成時に、答弁原文を加味することはむしろ悪影響を及ぼすと推測される。一方で、答弁要約生成時には、質問原文も答弁原文も与えた方が、答弁要約内容の精度が向上した。以上から、時系列の流れに沿って原文を入力することが最善であると考察する。

また、現在のモデルは、質問原文 1 つと答弁原文 1 つのペアに対して学習を行うが、実際の質疑応答では、質問原文 1 つに対して答弁原文が複数あることがある。この場合、質問要約生成時に答弁原文を加味すると、質問原文に対して質問要約が一意に定まらない。このことから、質問要約生成時に答弁原文を加味すべきではない。

5.4 カスケード型生成器の導入

モデル 4(.SET_AQPS) とモデル 5(.SET_cascade_AQPS) を比較し、カスケード型生成器の有効性について検証する (表 7)。カスケード型生成器を導入することで、すべての評価指標が向上した。A-content 指標のみならず Q-Content 指標も向上していることから、答弁要約の品質が向上すると、GAN の学習の中で、質問要約の品質も向上するよう、間接的に学習圧がかかっている可能性がある。

5.5 原文と要約の対応を判定する識別器 D_{QPS} , D_{APS} の導入 (AQPS)

モデル3(_SET_cascade)とモデル5(_SET_cascade-AQPS)を比較し、AQPSの有効性について検証する(表7)。AQPSを導入することで、すべての評価観点において良い結果が得られた。AQPSの導入により、生成要約内容の品質が向上し、これに伴って Dialog-goodness, Overall-goodness 指標も向上したと思われる。

5.6 GAN によって何が得られたのか

GANの導入によって、質問答弁要約ペアの「流暢さ」が向上すると予想した(3.2.2節)。これを踏まえて、GANを導入することで何が得られたのか、定性的な判断も含めて検討する(表10に例を示す)。

まず、流暢さの1つの側面として、文法の正しさが考えられるが、well-formed 指標は、GAN適用・不適用の間で、ほとんど差が無い。もともと、事前学習済T5モデルが、洗練された文法構造を出力できる言語モデルだからであろう。

GANの適用により、特に政治課題の議論において頻繁に見かけられるような表現や語句を出力するようになり、政治トピックの要約として自然になる印象がある。また、GANを適用すると、質問側と答弁側で対応関係のある単語や表現を利用することが増える。

GANを適用していないモデル1(ベースライン)の出力については、質問要約と答弁要約のどちらを先に生成したのか解らず、読みづらい印象がある。GANを利用すると、この点が解消し、質問要約・答弁要約の順に読むと理解しやすくなる印象がある。これは、単語や表現の対応に起因する部分が大きいかもしれない。

以上より、「読みやすさ」とも言うべき要素が向上したと考える。「読みやすさ」は流暢さの1つの指標であると思われるので、我々の予想は正しかったと言える。

6 おわりに

事前学習モデルT5を生成型要約モデルとして訓練することで、議会会議録の要約文章生成を行った。ここで、質問答弁の対話関係を加味するために、GANを利用する新規手法を提案した。また、GANの訓練において、生成器Gの損失関数について、教師強制を行いつつ、識別器Dの損失を加味するような新たな枠組みの損失関数を導入した。この枠組みは、両モデルの入出力の形式が大きく異なり、独立に運用するしかない場合でも敵対的に訓練できる点に強みがある。様々な応用に可能性を見込める手法と考える。

さらに3つの追加の新規手法、すなわちカスケード型生成器、各生成器に質問原文と答弁原文の両者を入力する手法、原文と要約の対応を判定する識別器 D_{QPS} , D_{APS} の導入は、いずれも性能向上に寄与することがわかった。総合的には自動評価・人手評価ともに良い結果を得られるモデルを構成できた。特に、カスケード型生成器は、対話関係を加味しつつ要約を行うための容易な工夫である。議会会議録の要約生成に限らず、対話関係の見られるテキストに対する要約生成において、応用の効く手法であると考ええる。

今後の課題としては、以下が挙げられる。まず、5つのモデルを対象に人手評価を行ったが、GANや各種手法の寄与についてより詳細に調査するため、追加の実験を実施すべきである。さらに、訓練したモデルは、年度や個数などの数字を扱うことを苦手としている。最後に、学習や評価で扱う議会録データセットは、人手要約と原文の内容が対応していない場合があるなど、看過できない問題点があることが挙げられる。他の地方議会会議録についても利用しながら、この課題に対する解決策を探るべきと考える。

参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc., 2014.
- [2] Kazuma Kadowaki. Jirid at the ntcir-15 qa lab-poliinfo-2 task: An abstractive dialog summarization system for japanese assembly minutes. *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, Dec 2020.
- [3] Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, and Hongyan Li. Generative adversarial network for abstractive text summarization. *CoRR*, Vol. abs/1711.09357, , 2017.
- [4] Yasuhiro Ogawa, Yuta Ikari, Takahiro Komamizu, and Katsuhiko Toyama. Nukl at the ntcir-15 qa lab-poliinfo-2 task. *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, Dec 2020.
- [5] Hiromu Onogi, Kiichi Kondo, Younghun Lim, Xinnan Shen, Madoka Ishioroshi, Hideyuki Shibuki, Tatsunori Mori, and Noriko Kando. Forst: A challenge to the ntcir-15 qa lab-poliinfo-2 task. *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, Dec 2020.
- [6] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *CoRR*, Vol. abs/1910.10683, , 2019.
- [7] 木村泰知, 関根聡, 乾健太郎. 地方議会会議録の要約に向けて. 言語処理学会 第24回年次大会 発表論文集 (NLP2018), Vol. P5-3, pp. 596-599, Mar 2018.

単語集合の共通点の発見練習による発想能力の獲得支援システム

Support System for Acquiring Ideation Skills through Practice in Discovering Commonalities among Sets of Words

神部 尚樹^{1*} 砂山 渡¹ 服部 峻¹
Naoki Kambe¹ Wataru Sunayama¹ Shun Hattori¹

¹ 滋賀県立大学 工学部

¹ School of Engineering, The University of Shiga Prefecture

Abstract: 我々の創造的な社会活動においては、発想力を欠かすことができない。発想は過去の経験や知識をもとに、ある物事に対する関連事項を広く想起することから始まり、複数の物事の関連事項を結びつける、すなわち複数の物事の共通点を探る能力と強く関わると考えられる。そこで本研究においては、複数の単語の共通点を繰り返し考えることで、発想能力を鍛える練習ができるシステムを提案する。評価実験の結果、システムを用いた被験者の発想能力が鍛えられる可能性を確認した。

1 はじめに

近年、IT化や少子高齢化、グローバル化などが進み、人工知能による雇用の代替や人々の価値観の変化が起きている。オックスフォード大学の調査結果によると、20年後には日本の労働人口の半数が就いている職業がAIやロボットに代替されるとも推計されている。

しかしAIは現在、特定の分野やタスクに特化しており、自律して動的な判断や行動ができるレベルには届いていない。また人間の感情や物事の背景、理由を考慮することできない。ただ先に述べたようにテクノロジーによって仕事が代用されるようになってきており、自らで適切に取捨選択し活用する思考力や、革新的なアイデアを生み出す発想力などの付加価値の高くクリエイティブな能力の重要性は近年高まっている。

社会の変化に伴って思考力や発想力がより重要になりつつあるため、発想のプロセスを理解し、能力獲得を支援することに大きな意味がある。この共通点を確認する練習を繰り返すことで発想能力の獲得を支援し、AIなどの機械には真似できない思考力や発想力の向上の支援をすることを目指す。発想力を鍛えることで、他人と差別化でき自分の市場価値を高めることができる。

そこで本研究では発想能力の向上支援を行うことを研究目的として、「共通点の発見練習」を行うことで発想能力の獲得支援を実現する。

2 関連研究

2.1 スキル獲得支援に関する従来研究

将棋初学者に対局の中で指し手の候補を提示することで、ユーザの楽しさや興味を向上させることができるかに関する研究がある[1]。人間の行動をシステムにより支援することで人間の感情にどのように影響するかについて研究しており、本研究では人間の思考能力にどのように影響するかについて注目する。

また以前に自身が経験した事柄を、別の事柄に当てはめて喩えることで理解を促す「類推による説明」を支援するシステムに関する研究がある[2]。特定の2つのものの関係を整理して説明スキルを鍛える研究で、これは類推の手順の理解を促すなかで共通点の認識を行っているがその能力を鍛えることを目的としているわけではない。

2.2 発想のトレーニングに関する従来研究

言葉と言葉を関連付けを行うことで発想力を向上させる研究がある[3]。発想力を4種類の能力に分解できると考え、数分間のトレーニングを複数日で行うことで発想力の向上ができるかという研究で、本研究では発想力向上の支援に関して自動化を行うことに新規性がある。特定の1つのキーワードと任意のキーワードの組合せを提示して新たな発想を支援する研究[4]がある。それらの組み合わせを数秒ごとに切り替えることで新たな発想を生み出す支援を行っており、発想の幅を広げることを支援している。

*連絡先： 滋賀県立大学工学部 電子システム工学科
神部 尚樹
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: on23nkambe@ec.usp.ac.jp

このようにシステムが発想を行う研究や、システムが人間の発想を助ける研究はあるが、人間の発想能力自体に注目しスキルの獲得支援を行う研究はみられない。

3 共通点発見練習システム

3.1 発想と共通点の発見練習

3.1.1 発想とは

発想とは単に思いつきではなく、過去の経験や知識を基に新たなアイデアを生み出すことで、過去の偉大な学者だけが起こすような可能な壮大なものではなく、誰でも頭の中で起こすことができる。

ある一定の材料となる知識や経験を持っていることを前提として、頭の中の知識ネットワーク上の関連する知識間に関連付けを与えることで行われている。さらに発想は大きく3つの形態に分類して定義されている。

- 「連想」：以前とほとんど同じ問題や状況に対して以前とほとんど同じ解決策を思いつくこと
- 「転用」：以前とは大分異なる問題や状況に対して以前とほとんど同じ解決策を思いつくこと
- 「独創」：以前とは大分異なる問題や状況に対して以前とは異なる解決策を思いつくこと

これら連想、転用、独創は自分が保有している知識を他方に転移する際にモデル(抽象)化を行い共通点を認識している。抽象化を行うことで共通点を発見する練習を繰り返し発想能力を鍛えることができ、本研究では「連想」と「転用」の基本的な発想の形態に注目し、スキル獲得を支援することを目的とする。

3.1.2 共通点の発見練習

一般にスキル獲得のためには簡潔に要素を抜き出したシステムでの練習の繰り返しを行う必要がある。複数の単語に関する情報をそれぞれ想起することで、その中から共通点を見つける練習を行うことで連想の能力を鍛えることができる。そこで共通点を発見し発想能力を鍛える手段として本研究では、表示された複数のキーワードと関連度の高い共通点となる単語の解答を繰り返すシステムを提案する。

図1に例として食べ物に関する単語の関連図を示す。

「りんご」や「ブドウ」というそれぞれのキーワードは同じ「果物」に属する同位語、それに対してまとめる「果物」をここでは上位語と定義する。さらに、りんごとトマトは食べ物という上位語に属するだけでなく、「赤い」、「丸い」などの単語とも関連度が高いとい

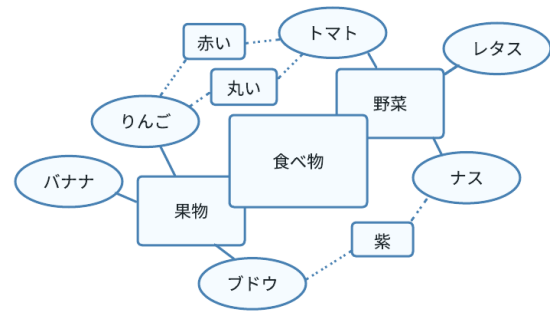


図1: 食べ物に関する関連単語の例

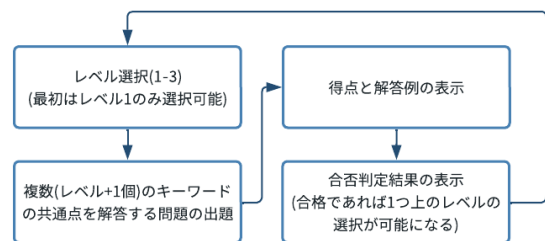


図2: 共通点発見練習システムの構成

える。本システムでは、ある複数の単語からその上位語や関連する単語となるものを考え共通点を発見、認識する練習を行うことで発想力向上の支援を行う。

3.2 共通点発見練習システムの構成

図2に共通点発見練習システムの構成を示す。本システムの利用の際、ユーザはまず出題レベルを選択する(図2左上)。レベルは3つあり、1つのレベルをクリアするごとに次のレベルに挑戦できるようになる。続いて、システムは選択されたレベルの共通点を発見する問題の出題(1セット10問)を行う。ユーザの解答後、得点と解答例を表示する。システムはユーザの得点をもとに合否判定を行い、合格であれば、ユーザは次のレベルの問題を選択できるようになる。不合格であれば、同じレベルの問題に再挑戦する。

3.3 レベル選択

キーワードが多いほどそれぞれと高い類似度をもつ単語の数は少なくなるように、共通点を発見する難易度は共通点を探す個数で決まる。よってレベル1をキーワード2つ、レベル2をキーワード3つ、レベル3をキーワード4つの出題に設定する。

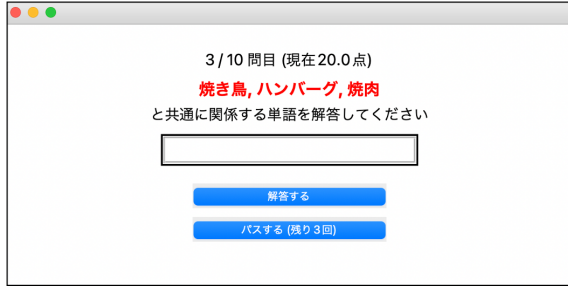


図 3: レベル 2 の共通点発見問題の出題画面

3.4 複数のキーワードの共通点を解答する問題の出題

図 3 にレベル 2 の共通点発見問題の出題画面を示す。ユーザは赤字の「焼き鳥」、「ハンバーグ」、「焼肉」の 3 つのキーワードに共通に関係する単語を解答する。ユーザは黒枠内に自由記述形式で単語を入力し「解答する」ボタンを押すことで解答する。対応する単語と単語ベクトルの組み合わせが存在すれば、それに応じた得点が表示され、そうでなければ違う単語を入力するように表示する。また 1 セットにつき 3 回まで「パスする」ボタンを押すことでキーワードを入れ替えることができる。これは得点や問題数の増減には影響しない。

問題の作成は fastText[5] の日本語学習済みモデルを用いた単語を数値化する分散表現を使用する。分散表現とは各単語を固定長のベクトル成分 (本システムで用いるモデルは 300 次元) で表すことで、ベクトル空間上でのベクトルとして表現することで単語ベクトル同士の角度による $\cos$ 類似度を求めることができ、単語同士の演算が可能になる。(1) 式に単語 $A = \{a_i (i = 1..300)\}$ と単語 $B = \{b_i (i = 1..300)\}$ の $\cos$ 類似度を示す。

$$\text{Cos_Similarity}(a, b) = \frac{\sum_{i=1}^{300} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{300} a_i^2} \sqrt{\sum_{i=1}^{300} b_i^2}} \quad (1)$$

図 3 に赤字で表示されるキーワードを漢字、ひらがな、カタカナで構成される名詞に限定する。wikipedia で学習した単語集合から 1 単語を想定する正解としてランダムで選択を行い、その単語との $\cos$ 類似度が 0.5 以上の単語をレベルに応じて必要数選択しキーワードとして設定する。表 1 に上記システムで生成されたキーワードとその解答例を示す。表示されるキーワードの組み合わせは必ずしも同位語や上位語などの関係はなく、ことばの持つ意味に応じて関連度が近いものを設定する。

表 1: システムで生成される問題と想定している解答例

表示されるキーワード	解答例
陽明学, 五経, 思想	朱子学, 儒教など
レーダー, 捕捉, ホーミング	ミサイル, 追尾など
貸切, タクシー, 運行	バス, 乗合など
正教会, カトリック エキュメニカル	キリスト教 プロテスタントなど
被ばく, 照射, シーベルト	放射線, 線量など
挽き, 穀物, 精白	大麦, トウモロコシ 大豆など
インスピレーション ヒント, アイデア	着想, 思想など
予備校, 志望校, 留年	受験, 入試など
SMAP × SMAP アンビリバボー アメトーク	特番, バラエティ 番組など

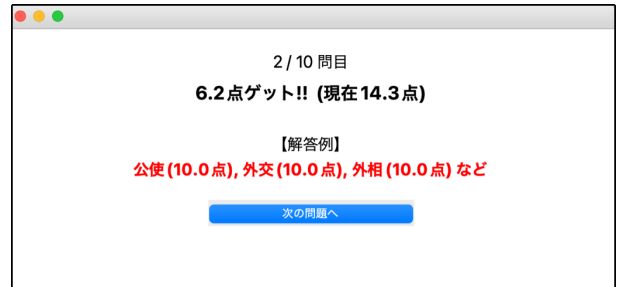


図 4: 共通点発見問題の解答結果の表示画面

3.5 得点と解答例の表示

図 4 に共通点発見問題の解答結果表示画面を示す。ユーザが解答する単語は表示されたそれぞれのキーワードとの関連性が求めるため、各キーワードとの $\cos$ 類似度の積に応じて 10 点を上限とする得点を与える。表 2 にその値と得点の関係を示す。全てのキーワードとの類似度が 0.55 以上であれば 10 点、0.33 以下であれば 0 点となる値を基準とする。すなわち各キーワードとユーザの解答単語の類似度が 0.55 から 0.33 の類似度であれば、式 (2) に示す近似された一次関数によって小数第 1 位までの得点を与える。また解答例として得点が 9 点以上の単語を 3 つ提示することで考え方の学習を促す。「次の問題へ」のボタンを押すことで次の問題の出題画面へと進み、10 問の解答と学習が終了すると合否判定結果表示画面へ進む。

$$\text{Score} = 47.05 \times \text{Cos_Similarity} - 4.23 \quad (n = 1)$$

$$\text{Score} = 71.09 \times \text{Cos_Similarity} - 1.94 \quad (n = 2)$$

$$\text{Score} = 119.9 \times \text{Cos_Similarity} - 0.97 \quad (n = 3) \quad (2)$$

表 2: 類似度と得点の関係

得点	0	...	10
類似度の積	0.30	...	0.55
2 類似度の積	0.090	...	0.30
3 類似度の積	0.027	...	0.16
4 類似度の積	0.0081	...	0.091

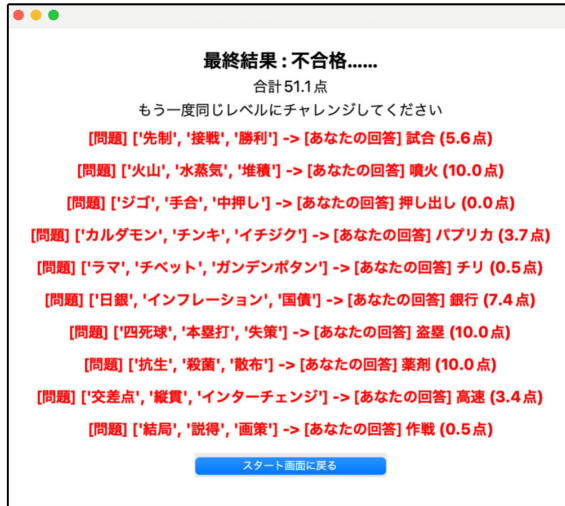


図 5: 共通点発見問題の最終結果の確認画面

3.6 合否判定結果の表示

図 5 に共通点発見問題の最終結果の確認画面を示す。これまでに行った 10 問の問題とそれに対するユーザの解答単語、それに対する得点、またクリア条件を満たしていれば「合格」、満たしていなければ「不合格」を表示する。「スタート画面に戻る」のボタンを押すことでスタート画面に戻り合格であればそのレベルに対応した数のキーワードの共通点を発見する能力が身についたため次のレベルへ進むことができ、不合格であればキーワードの共通点を捉える能力が不十分であるため再度同じレベルに挑戦を行う。

4 評価実験

本章では、本システムを使用することによって発想能力の獲得支援ができるかを検証する評価実験について述べる。

4.1 実験方法

10 代から 20 代の学生 8 名にシステムの使用前後にテストを行うことで提案するシステムの利用で発想能力を鍛える効果があるかを検証する実験を行った。実験手順は 3 つある。

表 3: 事前テストで表示するキーワード

	事前テスト
第 1 問	先制, 接戦, 勝利
第 2 問	火山, 水蒸気, 堆積
第 3 問	腹違い, いとこ, 孫娘
第 4 問	堤防, 空堀, 盛り土
第 5 問	印欧語, ドラヴィダ, サンスクリット
第 6 問	日銀, インフレーション, 国債
第 7 問	四死球, 本塁打, 失策
第 8 問	抗生, 殺菌, 散布
第 9 問	交差点, 縦貫, インターチェンジ
第 10 問	結局, 説得, 画策

- 実験手順 1 では、共通点を発見する「事前テスト」を行ってもらった。
- 実験手順 2 では、期間を 7 日間の間に 3 日間、本システムを 1 日最低 5 セット行ってもらい、レベル 1 からレベル 3 まで全てのレベルをクリアを目指してもらった。レベル 3 をクリア次第実験手順 3 を行ってもらった。
- 実験手順 3 では、再び実験手順 1 とは別の共通点を発見する「事後テスト」を行ってもらった。

4.1.1 事前テスト

はじめにレベル 2 の学習用問題から無作為に選出した 10 問をテストを行ってもらった。ただし事前テストが学習にならないように解答後に解答例は示さず、10 問のテストの終了後に得点の表示のみを行った。事前テストの内容を表 3 に示す。

4.1.2 システムによる学習

その後、システムを用いて共通点を発見する練習を行ってもらった。時間制限は問わずレベル 1 からレベル 3 までの合格を目指してもらった。各被検者は 1 日に最低 5 セットを 3 日間行ってもらった。

4.1.3 事後テスト

3 日間の学習を行ったのち、レベル 2 の学習用問題から無作為に選出した 10 問でテストを行った。事前テストと同様、解答後に解答例は示さず得点の表示は 10 問のテストの終了後にのみ行った。事後テストの内容を表 4 に示す。

表 4: 事後テストで表示するキーワード

	事後テスト
第 1 問	タブレット, モバイル, 端末
第 2 問	敗走, 士卒, 討ち死に
第 3 問	エジプト, イラク, イスラエル
第 4 問	駆逐, 艦艇, オルモック
第 5 問	野菜, 穀物, 綿花
第 6 問	実力, 新進, 才能
第 7 問	批判, リベラル, 擁護
第 8 問	序文, 原典, 校訂
第 9 問	力士, 土俵入り, 立合い
第 10 問	寵愛, 帰順, 側近

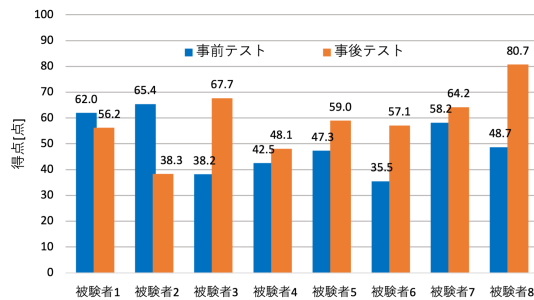


図 6: 各被験者ごとの事前テストと事後テストの得点

4.2 実験結果と考察

図 6 に各被験者ごとの事前テストと事後テストの得点を示す。

6 人の被験者の事後テストを比較して得点が高くなった。また 55 点未満の人数が事前テストでは 5 人に対して、事後テストでは 55 点未満は 2 人に減少している。これらのことから、本システムを通じた一定の学習の効果があったものと確認できる。

一方で、被験者 1 と被験者 2 は事後テストの得点が低くなった。この 2 名の事前テストの得点が高かったことから、学習前から一定の発想能力があり十分な学習効果が得られにくかったことが考えられる。

表 5 に被験者 3 が事前テストと事後テストで解答した単語とその得点を示す。事前テストに比べて、事後テストでは多くの問題で点数を獲得できるようになっていた。一方で、「艦艇」と「艦隊」、「穀物」と「作物」のように特定の単語の類義語や、「土俵入り」と「土俵」のように問題の一部の単語の解答で満点が与えられている事例も見受けられた。そのため、点数の与え方については、さらなる改良の必要性があると考えられる。

図 7 に各被験者ごとのレベル 2 における得点推移を示す。ほとんどの被験者が学習回数を重ねることによって得点が右肩上がりに増加している。一方で被験者 6

表 5: 被験者 3 の事前テストと事後テストの解答と得点

	事前テスト (得点)		事後テスト (得点)	
第 1 問	シーソーゲーム	2.2	PC	6.6
第 2 問	噴火	10.0	切腹	5.0
第 3 問	親戚	10.0	サウジアラビア	10.0
第 4 問	土手	10.0	艦隊	10.0
第 5 問	古代	3.4	作物	10.0
第 6 問	お金	0.4	センス	1.4
第 7 問	野球	0.1	評価	0.9
第 8 問	薬品	1.9	書物	9.6
第 9 問	車	0.0	土俵	10.0
第 10 問	計画	0.2	帰服	4.2
合計点	38.2		67.7	

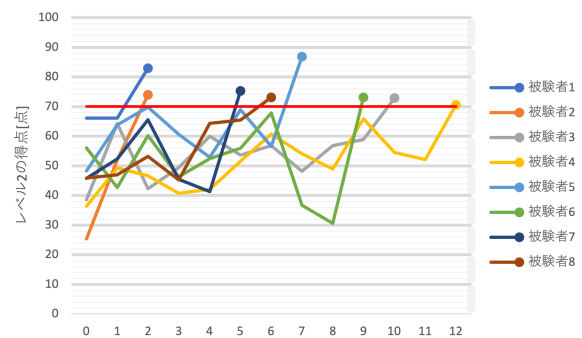


図 7: レベル 2 の得点推移

や被験者 7 が学習途中で点数が大きく下がっている。これは長時間の学習による集中力の低下だけでなく、問題がランダムであるため難しい問題が時折発生したためと考えられる。今回は実験の都合上、短期間で学習を実施してもらったが、実際にはより長期的に繰り返し練習してもらうことを想定しており、大きな問題にはならないと考えている。

学習回数が最も多かった被験者 4 は、図 7 において学習セット数の増加に伴って得点も上昇しており、事後テストの得点も事前テストより増加していることから、学習の効果は得られているものと考えられる。一方で、事後テストの点数自体は 50 点を下回っていることから、学習の速度には個人差があり、スキルを獲得してもらうためにはある程度時間をかけながら繰り返し練習してもらうことが必要と考えられる。

事前テストの点数が高かった被験者 1、被験者 2 については、レベル 2 を 2 回で合格しており、事前に一定の能力が備わっていることがうかがえるが、本システムによる学習を重ねられてはいないため、多様な問題には即座には対応できない可能性も考えられる。そのため、現在は 1 回の合格で次のレベルに進める形となっているが、しっかり学習してもらうためには、複数回の合格により次のレベルに進める形にすることも

表 6: 問題の難易度推定

難易度	問題数	確率	予想点	期待値
5	17	0.17	1	0.17
4	26	0.26	3	0.78
3	21	0.21	5	1.05
2	22	0.22	7	1.54
1	14	0.14	10	1.40

考えられる。

表 6 に示すように被験者が解いた問題 100 問を無作為で選出し、問題の難易度の推定を行った。難易度 5 から順に「とても難しい」「難しい」「普通」「易しい」「とても易しい」と設定し各問題がどの難易度に該当するか集計した。各難易度での期待値は確率と予想点の積であり、表 6 に示すようになった。また各難易度の期待値を合計した 1 問あたりで取れる点数の平均が 4.94 となった。よって 1 セット (10 問) に取れる点数の期待値が 49.4 点となるため、50 点以上の点数を取ることができれば発想能力は身につけていると考える。

4.3 アンケートによる考察

事後テストの後に被験者にアンケート調査を実施した。「学習を通して、複数のものの共通点を発見するスキルが向上したと思いますか?」という質問に対して 8 人中 7 人の被験者が「向上したと思う」と回答した。このことから、主観的にも本システムの学習による効果があるものと考えられる。

各レベルの難易度について「とても簡単」「簡単」「普通」「難しい」「とても難しい」の 5 択から回答してもらった。その結果、レベル 1 では「普通」、レベル 2 では「難しい」、レベル 3 では「とても難しい」と答えた被験者がそれぞれ一番多かった。このことから、レベル設定は概ね適切であったことがうかがえる。ただし、各レベルの中でも出題される単語によっては問題に難易度の差が発生しやすいことから、現在のレベルの中で、より詳細な難易度設定を行い、レベル数を増やすことで、より段階的に学習を行うことが効果的な可能性が考えられる。

4.4 システムの改善点

システムによる学習を繰り返すことで、単語の意味や類義語、対義語や上位語など関連のある単語を想像することができたという意見があった。一方で、特定のキーワードに注目しその類似語や対義語を入力するなど本来の目的に沿わない解答方法でも高い得点をとることができたり、上位語や下位語の判別がうまくで

きず、キーワードの共通点となる単語の判定が納得いかないという意見があった。また、意味がわからないキーワードが分からないままになっているので意味を解説してくれると良いという意見があった。

問題画面でわからないキーワードを調べることができる機能や、解答例における単語間の関係を表す例文を表示するなど、より考え方を習得しやすい学習支援環境にしていくことが望まれる。

5 おわりに

発想能力の獲得支援を目的として、複数のキーワードから共通点を発見することで発想能力の獲得を支援するシステムを実装した。このシステムを用いた学習による発想能力の獲得支援の可能性を見出せた。

今後は、より長期的な学習に対応できるように、キーワード数が同じレベルの中でも、キーワードによる問題の難易度の差によってレベルを詳細化することや、直近複数回の結果にもとづく合格ラインの設定、正解例の提示方法の工夫などによる改良を進めていきたい。

参考文献

- [1] 金澤聖也, 片上大輔: 好手提示に基づく将棋初学者支援システム, 2022 年度人工知能学会全国大会 (第 36 回), 4I1-OS-26a-01 (2022)
- [2] 砂山渡, 石田純太, 川本佳代, 西原陽子: 類推による説明スキルの獲得支援システム, 情報処理学会論文誌, Vol.59, No.10, pp.1922 – 1931 (2018)
- [3] 丹羽時彦, 神沼靖子, 松田稔樹, 雄山真弓, 中永睦子, 安田貢一: 発想力向上の取り組みと分析 IS 人材の基礎力育成に向けて, 情報処理学会研究報告, Vol.2011-IS-115, No.14, pp.1–8 (2011)
- [4] 上山雅樹, 小林慶太, 小川大海, 奥谷雄介, 佐久間亮介: 発想支援システム「ひらめきエンジン」の紹介, 2018 年度人工知能学会全国大会 (第 32 回), 4Pin1-44 (2018)
- [5] fastText: (URL) <https://fasttext.cc/> (2023 年 1 月 30 日確認)

多様な指標を考慮したコーディネート系列の生成

Generation of fashion coordination sequences considering various factors

菊池 愛海¹ 尾崎 知伸^{1*}
Aimi Kikuchi¹ Tomonobu Ozaki¹

¹ 日本大学 文理学部

¹ College of Humanities and Sciences, Nihon University

Abstract: This study addresses the automatic generation of fashion coordination sequences. Specifically, by using techniques on word embedding, topic modeling, and submodular optimization, we first propose a framework for generating fashion capsule wardrobes (CWs) that reflect the user's preferences in addition to the quality of the outfits and the diversity of styles. We then perform the selection of a multi-day fashion coordination sequences from the CWs under some realistic constraints, considering continuity and the schedule of each day. The effectiveness of the proposed framework is assessed through evaluation experiments on real datasets.

1 はじめに

近年、洋服の多様化が進み、個人の嗜好や時間・場所・場面を考慮した洋服の着合わせ、すなわちコーディネートを楽しむことは身近なものとなっている。その際、理想的なコーディネートに必要な洋服をすべて購入、準備できるとは限らず、既に所持している洋服から工夫してコーディネートが組むことが一般的である。しかし、場面に応じた適切な洋服を選択するとともに、自身の好きなスタイルに寄せたコーディネートが組むためには、一定の知識や経験が必要とされ、必ずしも容易ではない。更に日常生活を考えると、直近の予定を考慮して複数日分のコーディネートを用意する必要もあり、問題を難しくしている。

これらの問題を解決するための一つの方向性として、カプセルワードローブ (Capsule Wardrobe, CW) が挙げられる。CW とは、着合せ (組み合わせ) の良い多様なコーディネートを多数組むことのできる小数の洋服集合を意味し、アパレル業界におけるブランド戦略や個人の自己表現のための有用な手段としても認識され、近年多くの注目を集めている。またファッションに関する知識やセンスが不足している利用者の支援を目的に、いくつかの CW 自動構築アルゴリズムが提案されている [1, 2, 3, 4, 5]。これらの CW 自動構築アルゴリズムを利用することで、コーディネート準備の簡略化が見込まれるが、実際には自身の好みを CW へと

反映すること、および日々の予定を考慮しながら (洋服集合である) CW から具体的なコーディネート系列を考えることは必ずしも容易ではない。

これらのことを背景に、本研究では、トップス・ボトムス・シューズの3アイテムを対象とし、CW に着目したコーディネート系列の自動生成に取り組む。具体的には、単語埋め込み、トピックモデル、最適化の各技術を利用し、着合わせの良さとスタイルの多様性に加え、利用者の好みを反映した CW を生成する。さらに CW から、いくつかの現実的な制約の元、連続性と各日の予定考慮した複数日分のコーディネート系列を生成する技術を提案する。

本論文の構成は以下のとおりである。2章で先行研究について言及する。3章で提案手法の詳細を説明した後、4章で評価実験を行う。最後に5章でまとめを行い、今後の課題を述べる。

2 関連研究

これまでに、ファッションコーディネートに関して様々な研究が行われている。

例えば文献 [6] では、ファッションアイテムやカテゴリ、スタイルをそれぞれ実数ベクトル化することで、コーディネートスタイルに適したファッションアイテムを検索する手法を提案している。また文献 [7] では、ファッション画像とテキストを同一空間に埋め込むことで、テキストを用いたファッションアイテムの検索を実現している。

*連絡先：日本大学文理学部情報科学科
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: ozaki.tomonobu@nihon-u.ac.jp

一方、CW の自動構築に関しても、いくつかの研究が報告されている。もっとも初期の研究の一つである [1] では、画像処理技術を用いたファッションアイテムに対する属性抽出を通じ、コーディネートに文書（属性集合）化した上で、トピックモデル [8] を用いて算出される互換性と多様性を基準に CW を構築する。これに対し文献 [2] では、互換性と多様性の他に視覚的類似性等を CW 構築基準に加えることを、文献 [3] では、大域的特徴（色、形など）に加えて局所的特徴（袖、ロゴなど）の双方を利用して互換性を推定することをそれぞれ提案している。加えて [5] では、購入記録を用いた嗜好のモデル化と体形等を考慮した評価モデルを構築することで、CW の個人化を行っている。また [4] では、トップスとボトムスを対象とした二部グラフに対する最適部分グラフ抽出問題として、CW 構築を実現している。

詳細は後述するが、本研究では文献 [6, 7] 等と同様、ファッションアイテムの実ベクトル化を行う。また CW 構築アルゴリズムとして [1] による手法を採用する。これらを組み合わせることで、利用者の嗜好を反映した CW 構築を実現する。

3 コーディネート系列の生成

3.1 問題の定式化

本研究では、トップス・ボトムス・シューズの 3 アイテムを対象とし、CW に着目したコーディネート系列の自動生成に取り組む。以下では、対象とする問題の定式化を行う。

トップスの集合を T 、ボトムスの集合を B 、シューズの集合を S と表記し、各集合に含まれるファッションアイテム $o \in T \cup B \cup S$ を、そのアイテムに付与されたタグや説明から得られる単語の集合 $D_o = \{d_1, \dots, d_{|D_o|}\}$ と同一視する。加えて、単語の全体集合を $W = \bigcup_{o \in T \cup B \cup S} D_o$ と表記する。一方、コーディネート $c = \langle t, b, s \rangle$ を、各要素からそれぞれ一つのアイテム $t \in T, b \in B, s \in S$ を選択することで構成されるタプルで表現する。またトップス、ボトムス、シューズの各集合 $T' \subseteq T, B' \subseteq B, S' \subseteq S$ に対し、構築可能なコーディネート全体の集合を

$$C(T', B', S') = \{\langle t, b, s \rangle \mid t \in T', b \in B', s \in S'\}$$

と表記する。以上の準備のもと、以下に示す手順を経て、長さ N のコーディネート系列を抽出する。

1. 利用者によって与えられるスタイルを表すキーワード $w \in W$ に適合するトップス・ボトムス・シューズの集合 $T_w \subseteq T, B_w \subseteq B, S_w \subseteq S$ を特定する。

2. 文献 [1] による手法を用い、適合ファッションアイテムの集合 T_w, B_w, S_w からカプセルワードローブ $CW_w = \langle T_w^{cw}, B_w^{cw}, S_w^{cw} \rangle$ ($T_w^{cw} \subseteq T_w, B_w^{cw} \subseteq B_w, S_w^{cw} \subseteq S_w$) を抽出する。

3. i 番目のコーディネート c_i に対する嗜好を表すキーワード $w_i \in W$ を与え、いくつかの現実的な制約の元、後述する評価関数を最大化する長さ N のコーディネート系列

$$CS_w^{\{w_1, \dots, w_N\}} = [c_1, \dots, c_N] \\ (c_i \in C(T_w^{cw}, B_w^{cw}, S_w^{cw}))$$

を生成する。

以下、それぞれの手順について詳細を説明する。

3.2 適合ファッションアイテムの抽出

スタイルを表すキーワード $w \in W$ に適合するファッションアイテムを特定するために、単語埋め込み技術を利用する。

まず、各アイテム o に対する BOW 表現をトランザクションとするデータベース $\{D_o \mid o \in T \cup B \cup S\}$ を対象に単語埋め込み技術 word2vec [9, 10] を適用することで、各単語 $w \in W$ のベクトル表現 $\vec{w}$ を獲得する。次いで、得られたベクトル表現を用い、構成する単語の平均ベクトルとしてアイテム o のベクトル表現 $\vec{o} = \frac{1}{|D_o|} \sum_{d \in D_o} \vec{d}$ を獲得する。

以上の準備の下、閾値 σ を利用者によって与えられる自然数とし、キーワード w に対するベクトル間のコサイン類似度 $\text{sim}(w, o) = \cos(\vec{w}, \vec{o})$ の各集合 $X \in \{T, B, S\}$ における上位 σ 件を、 w に適合するアイテムの集合 $X_w \subseteq X$ とする。

3.3 カプセルワードローブの構築

適合ファッションアイテム集合 X_w ($X \in \{T, B, S\}$) からのカプセルワードローブ $CW_w = \langle T_w^{cw}, B_w^{cw}, S_w^{cw} \rangle$ の生成には、文献 [1] の手法を用いる。具体的には、構築する CW に含まれるアイテムの数 θ を指定し、後述する互換性に関する評価値 $C(\cdot)$ と多様性に関する評価値 $V(\cdot)$ に関し、その和が最大となる CW^* 、すなわち

$$CW^* = \arg \max_{\substack{CW = \langle T', B', S' \rangle, \\ T' \subseteq T_w, |T'| = \theta, \\ B' \subseteq B_w, |B'| = \theta, \\ S' \subseteq S_w, |S'| = \theta}} (C(CW) + V(CW))$$

を CW_w として獲得する。

CW に関する評価は、コーパスとして与えられるコーディネート集合 $C = \{\langle t_1, b_1, s_1 \rangle, \dots, \langle t_{|C|}, b_{|C|}, s_{|C|} \rangle\}$

を対象に、ファッションアイテムの種別を接頭辞とする単語から構成される文書の集合

$$\bigcup_{c=\langle t,b,s \rangle \in C} bow_c \text{ where } bow_c = \bigcup_{x \in \{t,b,s\}} \{x+d \mid d \in D_x\}$$

(記号 + は単語の連結を行う演算子) を獲得し、トピックモデル [8] を適用すること得られる確率モデル P を用いる。

文書 (単語集合) bow の P における尤度 $c(bow)$ を用い、 $CW = \langle T', B', S' \rangle$ の互換性、すなわち着合せの良さに関する評価値は、生成可能なコーディネート

$$C(CW) = \sum_{c \in C(T', B', S')} c(bow_c)$$

と定義される。一方、 P におけるトピック集合 G と文書 bow におけるトピック $g \in G$ の確率を $p(g \mid bow)$ を用い、 CW の多様性に関する評価値は、コーディネートによるトピックのカバー、すなわち

$$V(CW) = \sum_{g \in G} \left(1 - \prod_{c \in C(T', B', S')} (1 - p(g \mid bow_c)) \right)$$

と定義される。

3.4 最適コーディネート系列の特定

カプセルワードローブ CW_w からのコーディネート系列 $CS_w^{\{w_1, \dots, w_N\}} = [c_1, \dots, c_N]$ の生成には、以下の 4 評価関数の総和を用いる。

$$\begin{aligned} C(CS_w^{\{w_1, \dots, w_N\}}) &= \sum_{i=1}^N c(bow_{c_i}) \\ V(CS_w^{\{w_1, \dots, w_N\}}) &= \frac{1}{|G|} \sum_{g \in G} \left(1 - \prod_{i=1}^N (1 - p(g \mid bow_{c_i})) \right) \\ D(CS_w^{\{w_1, \dots, w_N\}}) &= \sum_{i=1}^{N-1} dist(\vec{c}_i, \vec{c}_{i+1}) \\ M(CS_w^{\{w_1, \dots, w_N\}}) &= \sum_{i=1}^N dist(\vec{c}_i, \vec{w}_i)^{-1} \end{aligned}$$

評価関数 $C(\cdot)$ は、コーパスから構築された確率モデル P における尤度を用いて、系列に含まれるコーディネートの着合わせの良さ (互換性) を評価する。また $V(\cdot)$ は、コーディネート集合の多様性を評価する。

一方、残り 2 つの評価関数 $D(\cdot)$ と $M(\cdot)$ は、適合ファッションアイテムの抽出に利用した単語やアイテムに対するベクトル空間を利用する。コーディネート $c = \langle t, b, s \rangle$ に対するベクトル表現を、構成要素のベクトルの平均値、すなわち $\vec{c} = \frac{1}{3} (\vec{t} + \vec{b} + \vec{s})$ とする。ま

た、2 つのベクトル $\vec{v}_1, \vec{v}_2$ 間の距離を $dist(\vec{v}_1, \vec{v}_2)$ と表記する。これらを用い、連続性に関する評価関数 $D(\cdot)$ は、連続するコーディネート c_i と c_{i+1} の違いが大きい系列に対し高い評価を与える。一方、キーワードに対する適合性に関する評価関数 $M(\cdot)$ は、指定されたキーワード w_i とコーディネート c_i の距離が近い系列を高評価する。

4 評価実験

提案したファッションコーディネート系列生成手法の有効性を評価するために、IQON3000[11] データセットを用いて評価実験を行った。実験では、765 のコーディネートを対象とし、トップス、ボトムス、シューズ以外のアイテムを削除するとともに、各アイテム o に対する説明文を形態素解析することで単語集合 D_o を獲得した。また、スタイルを表すキーワードを $w \in \{\text{カジュアル, かわいい}\}$ 、適合ファッションアイテム集合に関するパラメタを $\sigma = 10$ 、 CW に含めるアイテム数を $\theta = 5$ 、生成する系列長を $N = 4$ とした。加えて、各日に対する嗜好を $[w_1, w_2, w_3, w_4] = [\text{かわいい, 冬, 軽やか, フォーマル}]$ とするとともに、以下の制約の下でコーディネート系列の生成を行った。

制約 1 : すべての日でトップスが異なる

制約 2 : 3 種以上のボトムスを含む

制約 3 : 尤度が一定値未満のコーディネートを含んではいけない

4.1 定量評価

実験では比較のため、合計評価値を最大化する系列に加え、各評価値 ($C(\cdot), V(\cdot), D(\cdot), M(\cdot)$) を最大化する系列およびいくつかの評価値の和を最大化する系列を特定した。実験結果を表 1 に示す。

表 1 より、 $w = \text{カジュアル}$ の場合、妥当性 (C) だけに着目してコーディネート系列を生成しても、他の基準での評価が低いものが生成されてしまうことが伺える。一方、多様性 (V) と連続性 (D) は関連が強いこともあり、多様性 (V) を最大化するコーディネート系列は、他の基準でも比較的高い評価が得られることが示された。また多様性を最大化するコーディネート系列と、全評価基準の総和を最大化するコーディネート系列とを比較した場合、特に適合性 (M) の評価に大きな差が認められ、複数の基準を同時に考慮することで、各評価においてバランスの良いコーディネート系列が生成されることが示唆された。

表 1: 実験結果：得られたコード系列の評価値

最大化基準	C	V	D	M	合計
$w = \text{カジュアル}$					
C	0.232	0.128	0.209	0.025	0.594
V	0.210	0.303	0.211	0.061	0.785
D	0.214	0.223	0.225	0.051	0.713
M	0.134	0.172	0.165	0.224	0.695
$C + V$	0.212	0.284	0.171	0.041	0.708
$D + M$	0.157	0.212	0.219	0.114	0.702
$C + V + D$	0.195	0.201	0.211	0.036	0.643
$C + V + M$	0.201	0.221	0.192	0.074	0.688
$C + V + D + M$	0.200	0.219	0.216	0.181	0.816
$w = \text{かわいい}$					
C	0.301	0.122	0.196	0.032	0.551
V	0.137	0.210	0.201	0.018	0.566
D	0.122	0.208	0.241	0.019	0.590
M	0.115	0.180	0.185	0.194	0.674
$C + V$	0.218	0.182	0.163	0.022	0.585
$D + M$	0.152	0.185	0.237	0.153	0.727
$C + V + D$	0.222	0.194	0.221	0.041	0.678
$C + V + M$	0.190	0.198	0.189	0.123	0.700
$C + V + D + M$	0.265	0.209	0.229	0.112	0.815

$w = \text{かわいい}$ の場合、適合性 (M) に着目した最大化が総合評価の高いコーディネート系列の生成に寄与することが伺える。また、全評価基準の総和を最大化するコーディネート系列における各基準での評価値は、単一基準を最大化する系列の値には及ばないものの、それぞれ十分に高い値を示しており、欠点の少ないコーディネート系列が生成されていると考えられる。

4.2 定性評価

定性評価として、服装に気を遣う 20~30 代の女性 20 名に対してアンケート調査を行った。アンケートでは、本研究の方法で生成したコーディネート系列に対し、以下の 3 項目の質問を提示した。

質問 1：コーディネートはまともですか？(1：まともではない ~ 4：まともである)

質問 2：毎日異なる印象を受けますか？(1：受けない ~ 4：受ける)

質問 3：1 日目~4 日目でそれぞれどのような印象を受けますか。以下のスタイルを当てはめてください(かわいい, 冬, 軽やか, フォーマル)

質問 1 は、コーディネートの妥当性に関する評価に対応するものであり、コーディネートが妥当かどうか

表 2: アンケート結果：質問 1・質問 2

w	1	2	3	4
質問 1				
カジュアル	14.3%	28.6%	38.1%	19.0%
かわいい	0.0%	14.3%	28.6%	57.1%
質問 2				
カジュアル	4.8%	0.0%	19.0%	76.2%
かわいい	19.0%	23.8%	38.1%	19.0%

表 3: アンケート結果：質問 3

選択肢	1 日目	2 日目	3 日目	4 日目
$w = \text{カジュアル}$				
かわいい	45%	0%	30%	0%
冬	5%	0%	5%	60%
軽やか	5%	65%	5%	0%
フォーマル	20%	0%	30%	25%
未回答・その他	25%	35%	30%	15%
$w = \text{かわいい}$				
かわいい	45%	10%	5%	0%
冬	15%	45%	0%	15%
軽やか	0%	20%	5%	55%
フォーマル	5%	0%	70%	0%
未回答・その他	35%	25%	20%	30%

を評価する。また質問 2 は、多様性に関する評価に対応するものであり、毎日なるべく異なるコーディネートが生成されているかを評価するものである。一方質問 3 は、キーワードを用いたコーディネートに対する評価・確認に対応する。正答率が高いほど視覚的にも良いと言える。回答者に提示するコーディネート系列には、各コードを構成する画像のみを示し、使用したキーワードの説明はすべて省略している。アンケートの集計結果を表 2 に示す。

表 2 より、まず $w = \text{カジュアル}$ の場合、質問 2 よりコーディネートの多様性では高い結果を得られた。しかし、質問 1 の妥当性においては評価が分散し、妥当性があるとは断言できない結果となった。表 3 よりキーワード通りの印象を受けているの評価に関しては、1 日目の「かわいい」のキーワードに関しては被験者も「かわいい」という印象を受けたことが分かる。しかし、その後の 2 日目から 4 日目に関してはキーワード通りの印象を受けていないことが分かり、表 2 の結果からも照らし合わせると、コーディネートとしては妥当だがキーワード通りの印象は受けていないという結果になった。

次に $w = \text{かわいい}$ の場合、質問 1 よりコーディネートの妥当性で一定の高い結果を得られた。質問 2 より、 $w = \text{カジュアル}$ に比べ、コーディネートの多様性で

は評価が分散し低い結果となった。表3よりキーワード通りの印象を受けているかの評価に関しては、1日目と2日目のキーワードに関しては被験者もキーワード通りの印象を受けたことが分かり、 $w = \text{カジュアル}$ よりも良い結果となった。しかし、その後の3日目と4日目に関してはキーワード通りの印象を受けていないことが分かる。表2の結果からも照らし合わせると、ある程度キーワード通りの印象を持つ妥当なコーディネート系列が生成されているという結果となった。

5 まとめと今後の課題

本研究では、カプセルワードローブに着目したコーディネート系列の生成手法を提案した。より具体的には、トップス・ボトムス・シューズからなるコーディネートを対象に、指定されたスタイルを反映したCWの構築手法、および互換性と多様性に加え、連続性や嗜好との適合性を考慮したコーディネート系列の評価関数を提案した。また実データを対象とした定量・定性評価により、その有効性を確認した。

今後の課題として、より多様なスタイル・嗜好を用いて生成される長大なコーディネート系列を対象とした提案手法の評価が挙げられる。また、より詳細かつ柔軟にスタイルや嗜好を反映する仕組みの開発や、流行や視覚的側面、コストなどを考慮した評価関数の提案が挙げられる。

謝辞 本研究の一部は、JSPS 科研費 JP22K12173 の助成を受けたものである。

参考文献

- [1] W.-L. Hsiao and K. Grauman : Creating Capsule Wardrobes from Fashion Images, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7161–7170, 2018.
- [2] Y. Tanaka and T. Ozaki : An extension of optimal fashion capsule wardrobe construction by considering visual dissimilarity and number of good coordinates, *Proc. of the 2022 Tenth International Symposium on Computing and Networking Workshops*, pp.224–228, 2022.
- [3] H. Chen, Y. Lin, F. Wang, and H. Yang : Tops, Bottoms, and Shoes: Building Capsule Wardrobes via Cross-Attention Tensor Network, *Proc. of the 15th ACM Conference on Recommender Systems*, pp.453–462, 2021.
- [4] S. Patil, D. Banerjee, and S. Sural : A Graph Theoretic Approach for Multi-Objective Budget Constrained Capsule Wardrobe Recommendation, *ACM Transactions on Information Systems*, 40(1):1–33, 2022.
- [5] X. Dong, X. Song, F. Feng, P. Jing, X.-S. Xu, and L. Nie : Personalized Capsule Wardrobe Creation with Garment and User Modeling, *Proc. of the 27th ACM International Conference on Multimedia*, pp.302–310, 2019.
- [6] 知見 優一, 北山 大輔 : ファッション SNS におけるアイテム・スタイル特徴に基づく検索システムとその評価, 第12回データ工学と情報マネジメントに関するフォーラム, B2-2, 2020.
- [7] A. Rubio, LongLong Yu, E. Simo-Serra, and F. Moreno-Noguer : Multi-modal joint embedding for fashion product retrieval, *Proc. of the 2017 IEEE International Conference on Image Processing*, pp.400–404, 2017.
- [8] G. Xun, Y. Li, W. X. Zhao, J. Gao, and A. Zhang : A Correlated Topic Model Using Word Embeddings, *Proc. of the 26th International Joint Conference on Artificial Intelligence*, pp.4207–4213, 2017.
- [9] T. Mikolov, K. Chen, G. Corrado and J. Dean : Efficient Estimation of Word Representations in Vector Space, arXiv preprint, arXiv:1301.3781, 2013.
- [10] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean : Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems* 26, pp.3111–3119, 2013.
- [11] X. Song, X. Han, Y. Li, J. Chen, X.-S. Xu, and L. Nie : GP-BPR: Personalized Compatibility Modeling for Clothing Matching, *Proc. of the 27th ACM International Conference on Multimedia*, pp.320–328, 2019.

攻撃的コメントの発信抑制に向けたオンラインコメントの受け手の精神ダメージ推定

Mental Damage Estimation of Online Comment Recipients to Suppress Sending Offensive Comment

伊藤 誠基^{1*} 砂山 渡² 服部 峻²
Masaki Ito¹ Wataru Sunayama² Shun Hattori²

¹ 滋賀県立大学大学院 工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学 工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: 近年、オンラインコミュニケーションが盛んに行われるようになるにつれ、炎上やネットいじめなどの問題が起こるようになってきている。このような問題を避けるためには、コメントの発信者が発信相手の状況を客観的に理解し、発信したコメントによる影響を考えられるようになることが大事と考えられる。そこで本研究では、コメントの受け手の精神ダメージの量を推定し可視化するとともに、そのダメージ量に応じた攻撃的コメントの発信を抑制するメッセージを提示できるシステムを提案する。実験により、システムが推定するダメージ量と、提示するメッセージによる攻撃的コメントの発信抑制効果との関係を検証した。

1 はじめに

近年、オンラインコミュニケーションが盛んに行われるようになるにつれ、炎上やネットいじめなどの問題が起こるようになってきており、特にSNS上での誹謗中傷等の深刻化が問題となっている[1]。インターネット上における誹謗中傷の場面においては、意図的もしくは無意識に攻撃的コメントが発信されることがある。既存のオンラインコミュニケーションツールの多くは、あらかじめ発信することが適切でない単語を用意して、その単語を含むコメントを発信できないようにしている場合が多い。しかし、あらかじめ想定していない単語が含まれるコメントは防ぐことができないため、例えば「死ね」という単語は検出できるが、「氏ね」という単語を検出できない。また、攻撃の意図を持ってしまった時点で、そのような禁止ワードを避けた表現を考えることは難しいため、インターネット上における誹謗中傷問題の根本的な解決につながらない。

このような問題の解決を図るためには、コメントの発信者が発信相手の状況を客観的に理解し、発信したコメントによる影響を考えられるようになることが大事と考えられる。「赤信号みんなで渡れば怖くない」の

ように、冷静に考えると避けるべき事柄であっても、周囲の状況によってはそれが見えなくなることがある。

そこで本研究では、インターネット上のコメントの受け手の精神ダメージの量を推定し可視化するとともに、攻撃的なコメントを発信しようとした際に、相手のダメージ量に応じた攻撃的コメントの発信を抑制するメッセージを提示することで、その攻撃的なコメントの発信を防止するシステムを提案する。

以下、本論文では、2章で関連研究について述べ、3章で攻撃的コメントの発信抑制システムについて述べる。4章で提案システムの効果を検証した実験について述べた上で、5章で本論文を締めくくる。

2 関連研究

2.1 オンライン投稿における攻撃的コメントの検出に関する研究

入力されたコメントに含まれる単語の悪口度を算出して、そのコメントが悪口文であるか自動判別する研究[2]や、コメントに含まれる単語がどれだけいじめと関連するかの程度を数値で表した、いじめ表現辞書を用いてTwitter上のネットいじめを検出する研究[3]がある。

*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻
伊藤 誠基
〒522-8533 滋賀県彦根市八坂町 2500
E-mail: ov23mito@ec.usp.ac.jp

これらの研究では、個々のコメントにおいて、人を傷つける単語に重みをつけ、入力されたコメントが人を傷つける程度を判定しているが、それ以前の過去のコメントの影響を考慮していない。本研究では、単一のコメントだけではなく、コメントの時系列データを対象として精神ダメージの推定を行う。

2.2 オンライン投稿におけるコメントの分析に関する研究

Word2Vec や fastText などの単語の分散表現を用いて、ニュースサイトのコメント分析し炎上を検出する研究 [4] や、YouTube のコメントのベクトル化によりコメントを分析して炎上を検出する研究 [5] がある。

これらの研究では、発信された全てのコメントに数値を与え、その値から炎上の程度、すなわちコメントの受け手をどのくらい傷つけるか評価しているが、ネガティブなコメントの量を評価しており、実際に人を傷つけるかどうかまで踏み込んではいない。本研究では、人を傷つける可能性がある単語とその程度を用いて、精神ダメージの推定を行う。

2.3 攻撃的コメントの投稿取り下げに効果的なメッセージに関する研究

ネットいじめに関わる攻撃的コメントの投稿を防ぐために、Web アンケート調査結果を分析し、相手の立場に立ち考えることを促すセルフトークメッセージの効果を検討した研究 [6] がある。この研究では、攻撃的コメントを発信する際に、受け手が攻撃的コメントによって傷つくことを共感させる「共感」、攻撃的コメントを発信しないことのメリットを提示する「メリット」、攻撃的コメントを発信することのデメリットを提示する「デメリット」、攻撃的コメントの発信を禁止する「禁止」の4種類のメッセージによって、攻撃的コメントの発信を抑制するメッセージを提示する。また、車の速度違反抑制に効果的なメッセージについて検討した研究 [7] においては、説得において反発心を喚起させにくい「感謝」のメッセージの効果が高いことが示されている。

本研究においては、これらの5種類のメッセージを発信を抑制するメッセージとして利用することに加えて、これらのメッセージが効果的となるタイミングについて検証を行う。

2.4 本研究における攻撃的コメント

Vangelisti らは、人が傷つく要因とは、自分との関係が軽視されること、屈辱を受けること、言語的な又

は非言語的な攻撃であること、変えることのできない欠陥に関すること、不適切、悪意が感じられるユーモアであること、誤解されること、落胆させられること、ショックを受けさせられることとまとめている [8]。そこで本研究では、非言語的な攻撃を除く、人の傷つく要因を満たす単語を「相手の存在を否定する単語、相手との関係を軽視する単語、物理的に危害を加えることを示唆する単語、人格や容姿を貶す単語、相手を貶める単語」と定義した上で、これらの単語を「攻撃単語」と定義する。

実際的には、単純に攻撃単語を含んでいても、対話相手を攻撃する意図を含まないコメントも存在する。たとえば、軽口が言い合える仲の良い友人同士で、少し失敗した相手に対して、「おまえバカだよな」とコメントしたり、「周りはブスばかりだけど、君は美人だね」のように、相手に対しては攻撃しないが、その他の人を攻撃する場合などが考えられる。

しかし前者については、どこまで相手が許してくれるか不透明な可能性があり、無意識的にでも傷つける可能性がある単語の使用は避けた方が良くと考えられること、また後者については、第三者への攻撃についても、見ていて気分を害する人がいる可能性や、そのような単語に触発された人が、攻撃単語を使ってしまう可能性なども考えられるため、本研究においては、攻撃単語の使用そのものを抑制の対象とする。

3 攻撃的コメントの発信抑制システム

本研究においては、複数のユーザがお互いにオンラインコミュニケーションを取る場面において、各ユーザが攻撃的コメントによってダメージが蓄積されることがない環境の構築を目指す。本章で述べる攻撃的コメントの発信抑制システムにおいては、一人のユーザが他のユーザに攻撃的コメントの投稿を行う場面に着目し、そのコメントの受け手となるユーザの精神ダメージを推定し、投稿を行おうとしているユーザの攻撃的コメントの発信を抑制する場面を想定している。

3.1 攻撃的コメント発信抑制システムの構成

図1に攻撃的コメントの発信抑制システム¹の構成を示す。

図中の「外部サイト」は、ユーザがオンラインでコミュニケーションを取るための環境を表す。あるユーザがそのサイト内で、別のユーザに対してコメントの

¹本システムの実装に際しては、テキストマイニングのための統合環境 TETDM(Total Environment for Text Data Mining)[11, 12]を利用した。

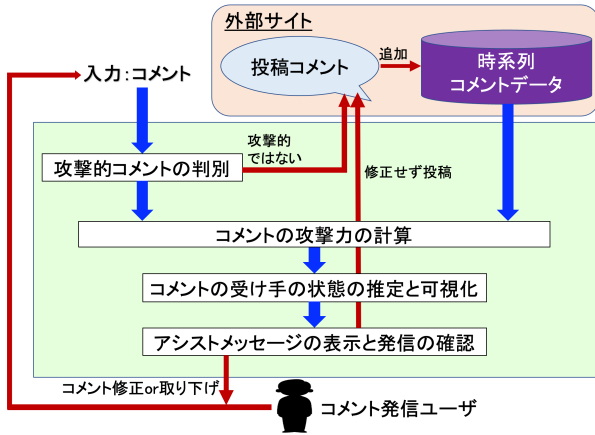


図 1: 攻撃的コメント発信抑制システムの構成

発信を行う場面において、発信しようとしているコメントをシステムへの入力とする。システムは、コメントが攻撃的コメントに該当するかどうかを判定し、攻撃的コメントでなければ、そのコメントは外部サイトに投稿される。攻撃的コメントとして判定された場合は、そのコメントを受信するユーザーがそれまでに受けたコメントによる蓄積ダメージを表す状態 (HP: 精神ダメージ許容量) を推定し、コメントの発信を抑制するアシストメッセージを、ユーザーにフィードバックしてコメントの修正や取り下げを促す。

3.2 攻撃的コメントの判別

ユーザーによって入力されたコメントに、以下で定義するいずれかの攻撃単語が含まれているコメントを「攻撃的コメント」として判定する。攻撃単語の一覧を表1に示す。これらの分類は、2.4 節の関連研究に基づいて行い、その攻撃の強さを3段階に分けた。これらに該当する単語は、日本語評価極性辞書 [9] においてネガティブと判定される単語の中から、上記分類に当てはまる単語を手動で選択するとともに、それらの単語の類義語を、類語辞典 [10] を用いて用意した。

3.3 コメントの攻撃力の計算

前節の単語の攻撃力に基づき、コメントの攻撃力の計算を行う。コメントの攻撃力は、あるユーザーが受けた時系列コメント全体による精神ダメージの推定量、ならびに投稿されようとしている攻撃的コメントが、これから与える精神ダメージの推定量に用いる。

コメントの攻撃力は、コメントに含まれる攻撃的単語の攻撃力の和で与える。ただし、単語 w の攻撃力 $atk(w)$ は、コメントを受け取る人の直近の 100 コメン

トにおける、単語 w の出現回数を $count(w)$ として、単語の基礎攻撃力 $base_atk(w)$ を用いて式 (1) により与える。これは、同一単語が複数回使われたときには、その精神ダメージとなる感覚量は対数に比例して減少する「ウェーバー・フェヒナーの法則」に基づいている。また式中の、 hp_w は後述のユーザーの精神ダメージ許容量 (HP) の最大値に応じた重みを表し、本研究では $hp_w = 14$ に設定している。

$$atk(w) = base_atk(w)(1 - \log_9 count(w)) \times hp_w \quad (1)$$

続いて、コメント c の攻撃力 $attack(c)$ は、コメント c に含まれる攻撃単語 w_i の攻撃力の和と、そのコメントの攻撃意図の程度 $atk_int(c)$ を用いて、式 (2) により与える。

$$attack(c) = \sum_i atk(w_i) \times atk_int(c) \quad (2)$$

ここで、コメント c の攻撃意図の程度 $atk_int(c)$ は、式 (3) で与える値として、0 から 70 の範囲で設定される値となる。

$$atk_int(c) = 0.01 \min\{-0.7atk_dis(c) + 70, 70\} \quad (3)$$

攻撃意図の程度 $atk_int(c)$ は、直前の攻撃的コメントとの間隔に基づいて、攻撃的コメントが連続する状況においては、攻撃的なコメントが発生しやすく、直前の攻撃的コメントからの距離が遠いほど、攻撃的なコメントが発生しにくいという傾向に基づいて定めた値である。この式は、YouTube Data API [13] を用いて収集した攻撃的コメントを多く含む炎上動画と、そうでない動画のコメントを各 4 万コメント (合計 8 万コメント) 収集した上で、それぞれで攻撃的コメントの発生間隔の分布を調べ、炎上動画のコメントを実際の攻撃意図を含むコメント、それ以外の動画のコメントを攻撃意図を含まないコメントとして算出した割合となっている。

たとえば、攻撃間隔 $atk_dis(c)$ が 0 のときの攻撃意図の程度 $atk_int(c)$ は 70%、攻撃間隔 $atk_dis(c)$ が 100 のときの攻撃意図の程度 $atk_int(c)$ は 0% となり、その他の間隔の値は両者を直線的に補間する値となる。

3.4 コメントの受け手の状態の推定と可視化

一連のコメントを受けているユーザーの状態を推定する値として、精神ダメージの許容量 (HP) を設定する。HP は 100 を初期値として、攻撃的コメントを受けるたびに、前節で定めたコメントの攻撃力の値が減少し、大きな精神ダメージにより、これ以上コミュニケーション

表 1: 攻撃単語の分類と点数

攻撃単語の分類	単語数	単語例	基礎攻撃力
・相手の存在を否定する単語 ・相手との関係を軽視する単語 ・物理的に危害を加えることを示唆する単語	48	「死ね」,「消えろ」,「殺す」 「邪魔」,「絶交」,「勘当」	3
・相手の人格を貶す単語 ・相手の容姿を貶す単語	306	「クズ」,「ケチ」,「意地汚い」 「ブス」,「デブ」,「ハゲ」	2
・上記以外の相手を貶める単語	303	「馬鹿」,「アホ」,「雑魚」 「ボケ」,「ドジ」,「低脳」	1

表 2: コミュニケーションの状態に応じた 1 コメントあたりの HP の回復量と最大 HP

状態	HP 回復量	最大 HP
攻撃モード	0.1	100
通常モード	0.5	110

ンを継続することができない状態を、HP が 0 の状態として設定する。

計算された HP は、次の 3 段階で色分けして、棒グラフで可視化する。

- 通常状態：HP が 71 以上、緑色で表示
- 注意状態：HP が 31 以上 70 以下の時、黄色で表示
- 危険状態：HP が 30 以下、赤色で表示

また、HP が減少するだけでは、いずれすべてのユーザの HP が 0 になってしまうことと、人は時間経過によってある程度ダメージを回復することができると考えられるため、攻撃的でないコメントが与えられるたびに、表 2 に示す量の HP を回復させる。ただし、精神ダメージの回復量は、攻撃的コメントが発生している状況と、そうでない状況とは異なると考えられるため、攻撃的コメントが発生している状況（攻撃モード）とそれ以外（通常モード）とで、1 コメントあたりの HP 回復量と回復後の最大 HP に差を設ける。

攻撃モードの判定条件は、直近の 100 コメントの中にある、攻撃コメント集合 ATK と攻撃コメントの数 $count_atk$ を用いて、式 (4) の条件を満たすときに攻撃モード、それ以外の時を通常モードとして判定する。すなわち直近の 100 コメントにおいて、式 (3) で計算される攻撃意図の程度が、平均 50 を超えた場合に攻撃モードと判定される。

$$\sum_{i \in ATK} atk_int(c_i) > 50 \times count_atk \quad (4)$$

これらの値は、Youtube の炎上動画のコメントとそうでない動画のコメントを元に、炎上動画では HP が 0 になり、それ以外の動画では HP が 100 付近で安定する値として設定した。

表 3: アシストメッセージとその種類

種類	アシストメッセージ
共感	逆の立場なら、あなたはそのコメントに傷つきませんか？
感謝	いつもありがとうございます。お互いに気持ちの良いやりとりができるようにご協力をお願いいたします
メリット	そのコメントを送らないことが、あなた自身の成長につながりますよ
デメリット	そのコメントを送ると、あなた自身の評価も下げることになりますよ
禁止	そのコメントを送ってはいけません

3.5 アシストメッセージの表示と発信の確認

本研究では、攻撃的コメントを発信しようとしているユーザの意識に変化を与え、攻撃的コメントの発信を抑制するメッセージをアシストメッセージと呼ぶ。

本研究で用意したアシストメッセージとその種類を、表 3 に示す。アシストメッセージには、2.3 節で述べた内容に基づいて攻撃的コメントにより、相手が傷つくことを「共感」させるメッセージ、「感謝」を伝えることで心理的リアクタンスを喚起させないメッセージ、攻撃的コメントを発信しないことによる「メリット」「デメリット」を伝えるメッセージ、および発信を「禁止」するメッセージを用意した。

これらのメッセージを表示するタイミングについて、次章で述べる検証結果に伴って設定する。このアシストメッセージを表示した後に、コメントを発信しようとしているユーザに対して、「本当にこのメッセージを送っても大丈夫ですか？」と確認を行うダイアログを表示させ、「はい」を選択した場合はコメントが投稿される。また、HP が 30 以下の場合には、「禁止」のメッセージを出して、コメントの発信を強制的に防ぐ。

3.6 攻撃的コメント発信抑制インタフェース

図 2 に、攻撃的コメントの発信抑制支援システムの画面を示す。

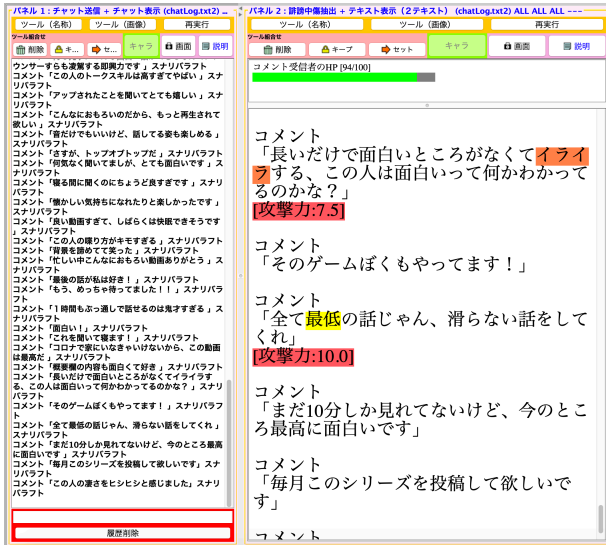


図 2: 攻撃的コメントの発信抑制支援システムの画面

左側のパネルの下部に、コメントを投稿するテキストフォームがある。右側のパネルの上部に、コメントを発信しようとしている相手の HP が表示され、その下に、そのユーザがそれまでに受けたコメントと攻撃単語が可視化して表示される。

攻撃単語は、基礎攻撃力 3 の単語が紫色、基礎攻撃力が 2 の単語がオレンジ色、基礎攻撃力が 1 の単語が黄色で表示される。またコメントの攻撃力も表示される。

4 攻撃的コメントの発信抑制システムの検証

本章では、提案する攻撃的コメントの発信抑制システムが提示する HP の妥当性とアシストメッセージの効果、ならびにアシストメッセージを表示するタイミングを検証するために行った実験について述べる。本検証は以下で述べるアンケート形式で 200 名の被験者に対して行った。被験者は 20 代、30 代、40 代、50 代の男女それぞれを 25 名ずつとした。また本調査は株式会社クロス・マーケティングに依頼して実施した。また、題材としてある動画投稿サイトに投稿された 7 件の動画に与えられた各 100 件のコメントをもとに実験を行ったが、実施に際して動画の投稿者やコメント発信者への影響を考慮して、動画のタイトルやコメントの内容は実験の本質に影響がない範囲で一定の改変を行って実施した。

表 4: 検証課題 1 に用いた動画のタイトルとシステムの推定最終 HP および投稿者のメンタルの強さ

動画のタイトル	HP	メンタル
ひたすら面白い話をする	94	普通
モデルのメイク紹介	78	普通
自分の動画の再生数があまり増えない	67	普通
お店のドライブスルーでドッキリをしにかけてみた	31	強い
いつも勝手に撮ってるけど自分を勝手に撮ってくる人に激怒してみた	22	強い
自身の炎上について歌ってみた	-41	強い

4.1 検証課題

4.1.1 課題 1: HP の妥当性の確認

提案システムが提示する HP の妥当性を確認するための課題を実施した。表 4 に示す 6 つの動画に与えられたコメント 100 件ずつを抽出し、そのコメントを読んでもらい、動画の投稿者がどの程度傷ついているかを、「特にダメージを受けていない」「少しダメージを受けているかもしれないが気にするほどではない」「多少ダメージを受けているがこれ以上耐えられないほどではない」「かなりダメージを受けていてこれ以上の傷つくコメントは見られない」の 4 択で回答してもらった。表中の HP は、各動画コメント 100 件を本システムに入力したときの最終の HP の値を示す。HP が 71 以上の動画を 2 つ、31 から 70 の範囲の動画を 2 つ、30 以下の動画を 2 つ用意して実施した。また動画の投稿者の立場で回答してもらうに際して、人によって精神ダメージへの強さが異なると考えられたため、動画の投稿者のメンタルの強さについて、著者の合議で「強い」と「普通」の切り分けを行った。

4.1.2 課題 2: 精神ダメージ量と効果的なアシストメッセージの種類の関係の確認

「街の女性に声をかけて勝手に容姿を評価してみた」というタイトルの動画に与えられた 100 件のコメントを用いて、被験者が「周りに迷惑をかけるようなアホなことをするな」とコメントを発信しようとしている状況を設定し、100 件のコメントによる精神ダメージと攻撃コメントを提示した上で、コメントの発信を取りやめるメッセージを、表 3 の 5 つと、「メッセージの内容によらず、このコメントを送る」「メッセージの内容によらず、このようなコメントを送る気にはならない」の中から選択して回答してもらった（複数回答可）。本課題は、一定の攻撃的コメントが発生しても止むを得ないと考えられる状況の中で、その攻撃的コメントの発信の取りやめに効果的なメッセージを探る目的で

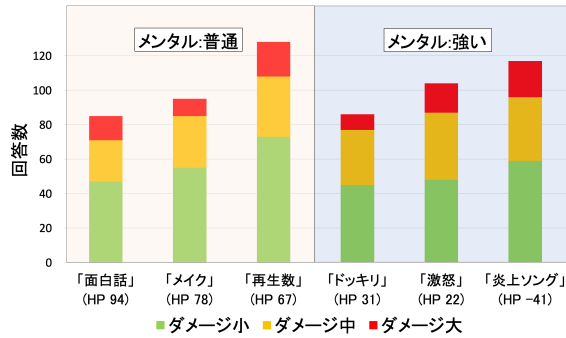


図 3: 各動画コメントに対して投稿者が傷つくと回答した人数

行った。提示する精神ダメージは、提示する攻撃コメントが発生したときに提案システムが算出する HP の減少量を用いて、0 から 100 まで 10 刻みで 11 パターンを用意し、ダメージ量を表す棒グラフとして提示した。一人の被験者にこの 11 パターンをランダムで提示し、各被験者には 11 回回答してもらった。

4.2 検証結果と考察

4.2.1 課題 1 の結果と考察

図 3 に、各動画コメントに対して投稿者が傷つくと回答した人数を示す。左半分のメンタルが普通の投稿者と、右半分のメンタルが強い投稿者のそれぞれにおいて、HP が減少するにつれ傷つくと判断した人の割合が多くなった。このことから、システムが推定する HP に一定の妥当性があることが確認できる。

一方で、前回答者数に対して、ダメージ中（多少ダメージを受けている）やダメージ大（かなりダメージを受けている）と回答した人数の割合は高くなく、また全く傷ついていないと回答した人が約半数（100 人前後）いた。これは、自分ではない他者の精神ダメージ量を過少に見積もってしまいやすいことや、それまでになされている 100 件（多量）のコメントを十分に読まずに勝手に大丈夫だろうと思ってしまうことなどが原因と考えられる。そのため、本システムによる精神ダメージの推定量を表す HP の提示は、攻撃的コメントの発信抑制に向けて、大きな意味を持つ可能性が考えられる。

4.2.2 課題 2 の結果と考察

表 5 に、メッセージの内容によらず、「このコメントを送る」または「このようなコメントを送る気にはならない」と回答した人数とその割合を示す。全体の 50%

表 5: メッセージの内容によらず、「このコメントを送る。」または「このようなコメントを送る気にはならない」と回答した人数と割合

ダメージ	送る	割合 (%)	送らない	割合 (%)
0	6	3.0	114	57.0
10	9	4.5	114	57.0
20	9	4.5	106	53.0
30	8	4.0	102	51.0
40	10	5.0	104	52.0
50	7	3.5	105	52.5
60	6	3.0	107	53.5
70	10	5.0	105	52.5
80	11	5.5	102	51.0
90	9	4.5	106	53.0
100	7	3.5	105	52.5

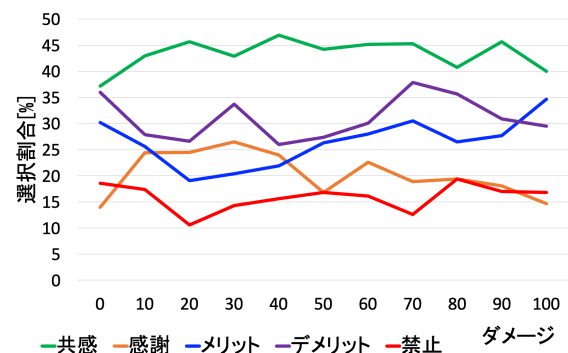


図 4: 攻撃的コメントの発信を取りやめるアシストメッセージの割合（攻撃的コメントの発信の意思がない人を除く）

強の被験者は、「このようなコメントを送る気にはならない」と回答したことから、約半数の人からは攻撃的コメントの発信の意思がないことが伺える反面、残りの半数弱の人は攻撃的コメントを発信してしまう可能性を示唆している。また全体の約 5% の人は、メッセージの内容によらず「このコメントを送る」と回答しており、これらの人に対しては提案システムによるアシストでは対応することができない。本研究における提案システムは、強い攻撃意志を持っている人ではなく、無意識やなんとなくで攻撃的コメントを発信してしまう人を対象としているが、今後このような人の一部にでも、意識の変化を促すことができる方法を模索することは重要と考えている。

図 4 に、表 5 で攻撃的コメントの発信の意思がなかった人について、攻撃的コメントの発信を取りやめるアシストメッセージの割合を示す。全ダメージ量において、相手の立場に立って考える「共感」の効果が最も高く、続いて「デメリット」の効果が高く、「禁止」が

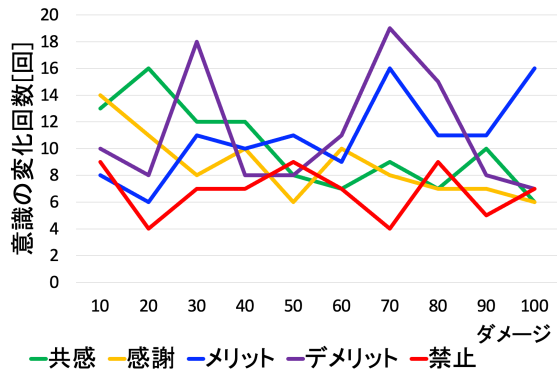


図 5: 攻撃的コメントの発信を取りやめる意識の変化の回数（ダメージ n-10 では攻撃的コメントの発信を行っていたが、アシストメッセージによってダメージ n でその発信を取りやめた人数）

最も効果が低い結果となった。「共感」「デメリット」の効果が高かったのは、間接的または直接的に自身がダメージを受けることを避ける保身的な考えが働いたためと考えられる。これは、「感謝」「メリット」を上回っていることから、日本人特有の自身が得をすることよりも、損をすることを嫌う傾向が反映されたものとも考えられる。また「禁止」の効果が低かったのは、理由もなく一方的に制限されることは、大きな心理的リアクタンスを生むためと考えられる。

図 5 に、各ダメージにおける攻撃的コメントの発信を取りやめる意識の変化があった回数（ダメージ n-10 では攻撃的コメントの発信を行っていたが、アシストメッセージによってダメージ n でその発信を取りやめた人数）を示す。

「共感」「感謝」のグラフは、いずれも右肩下がりのグラフとなっており、ダメージ量が少ない領域において、意識の変化をもたらしやすいことがわかる。これは、まだあまり攻撃的コメントが与えられていない状況においては、周りに流されることなく、相手の立場に立って客観的に物事を捉えたり、感謝の気持ちを受け入れやすかったためと考えられる。

また、「メリット」のグラフは右肩上がりとなっており、ダメージ量が多い領域において、意識の変化をもたらしやすいことがわかる。これは、一定の攻撃的コメントが与えられている状況においては、周りに流されやすい状況が発生しており、自分にメリットがなければ投稿を取りやめるきっかけにはなりにくいことを示唆している。そのため、先の「メッセージの内容によらず、このコメントを送る」と回答した人たちに対しても、より具体的に現実的なメリットを付与すること、たとえば、攻撃的でないコメントを送ることでポイントが付与されるなどの、強いメリットを提示

することが有効な可能性が考えられる。

「デメリット」のグラフは、左右にそれぞれ大きな山があり、デメリットによる意識の変化のタイミングは人によって 2 つのパターンに分かれたことが伺える。デメリットは、リスクとも捉えることができるため、リスクマネジメントにおいて、早い段階からリスクを避けようとする人と、一定のリスクを受容する人とに分かれた可能性が考えられる。

「禁止」は、ダメージ量によらず、意識の変化をもたらす回数は最も低い結果となった。

これらの結果、および図 4 においては全体的に「共感」のアシストメッセージの効果が高かったことから、以下のダメージ量に応じたアシストメッセージの提示方法が提案される。

- ダメージ量が少ない段階においては、まずは「共感」と「感謝」のアシストメッセージを提示し、続いて「デメリット」のメッセージを提示する。
- ダメージ量が中程度の段階においては、「共感」と「デメリット」のアシストメッセージを提示し、続いて「メリット」のメッセージを提示する。
- ダメージ量が多い段階においては、まずは「共感」と「メリット」のアシストメッセージを提示し、続いて「デメリット」のメッセージを提示する。

また、各種類のアシストメッセージとしては、表 3 に示したものの以外にもメッセージ例を用意することで、同一のメッセージによる飽きが来ないように配慮することも必要と考えられる。

5 おわりに

本論文では、インターネット上のコメントの受け手の精神ダメージの量を推定し可視化して、攻撃的なコメントを発信しようとした際に、相手のダメージ量に応じた攻撃的コメントの発信を抑制するメッセージを提示することで、その攻撃的なコメントの発信を防止するシステムを提案した。

検証実験の結果、システムが提示する精神ダメージの許容量（HP）が人間の感覚に近い一定の妥当性を持つことを確認した。また、5 種類のアシストメッセージについて、ダメージ量に応じた攻撃的コメントの発信抑制効果を検証した。その結果、ダメージ量が少ない場合においては、「共感」「感謝」のメッセージ、ダメージ量が多い場合においては「メリット」のメッセージが意識の変化をもたらしやすいことを確認した。

参考文献

- [1] 総務省：総務省におけるインターネット上の誹謗中傷対策の取組について, (URL) <http://urx3.nu/n5vv> (2023 年 2 月 16 日確認)
- [2] 石坂達也, 山本和英：Web 上の誹謗中傷を表す文の自動検出, 言語処理学会 第 17 回年次大会 発表論文集, pp.131 – 134 (2011)
- [3] 大友泰賀, 張建偉, 中島伸介, 李 琳：いじめ表現辞書を用いた Twitter 上のネットいじめの自動検出, 第 12 回データ工学と情報マネジメントに関するフォーラム (DEIM2020), C7-1 (2020)
- [4] Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Bruno Defude: Uncovering Flaming Events on News Media in Social Media, 2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC2019), pp.1 – 8 (2019)
- [5] 堀雄之介, 竹内幹太, 伊東栄典：コメントを利用した炎上動画検出に関する検討, 情報処理学会研究報告, Vol.2021-ICS-203, No.9, pp.1 – 5 (2021)
- [6] 西原陽子, 藤堂悠杜, 山西良典：セルフトークメッセージの呈示による悪口投稿取り下げの効果分析, 電子情報通信学会論文誌 A, Vol.J104-A, No.2, pp.19 – 26 (2021)
- [7] 五十嵐彩那, 臼井伸之介：速度違反抑制に効果的なメッセージと提示タイミング, 交通化学, Vol.46, No.1, pp.13 – 24 (2015)
- [8] Anita L. Vangelisti, Stacy L. Young, Katy E. Carpenter-Theune, and Alicia L. Alexander: Why Does It Hurt?: The Perceived Causes of Hurt Feelings, Communication Research, Vol.32, No.4, pp.443 – 477 (2005)
- [9] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一：意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203 – 222 (2005)
- [10] 類語辞典 weblio：(URL) <https://thesaurus.weblio.jp> (2023 年 2 月 16 日確認)
- [11] Total Environment for Text Data Mining (テキストデータマイニングのための統合環境)：(URL) <https://tetdm.jp> (2023 年 2 月 16 日確認)
- [12] 砂山渡, 高間康史, 徳永秀和, 串間宗夫, 西村和則, 松下光範, 北村侑也, 統合環境 TETDM を用いた社会実践, 人工知能学会論文誌, Vol.32, No.1, pp.NFC-A.1 – 12 (2017).
- [13] YouTube Data API: (URL) <https://developers.google.com/youtube/v3> (2023 年 2 月 16 日確認)

日常のSNS コメントに対する褒めリプライの自動生成システム Automatic Generation System of Compliment Replies to Daily SNS Comments

藤原 弘貴^{1*} 砂山 渡¹ 服部 峻¹
Koki Fujiwara¹ Wataru Sunayama¹ Shun Hattori¹

¹ 滋賀県立大学 工学部

¹ School of Engineering, The University of Shiga Prefecture

Abstract: 日本人は謙遜することが多く自分を褒めることが少ないため、自己肯定感を高めにくく、ストレスを溜めやすい状況がある。そのため、褒められる機会を増やすことは重要と考えられる。そこで本研究では、日常の出来事に関するSNS コメントに褒めリプライを自動生成するシステムを提案する。実験により、提案システムが作成する褒めリプライが、人を嬉しくさせる効果があることを確認した。

1 はじめに

現代社会はインターネット社会と言われている。インターネットを利用したオンラインコミュニケーションツールは非常に多く存在し、生活していく上でなくてはならないものへとになっている。

しかし、対面で会話をする場合と比較して文字でのコミュニケーションは感情が伝わりにくい。相手の表情や仕草、声の調子など感情が現れる部分を受け取れないため、文字のみの会話は無機質な言葉や攻撃的な言葉という印象を感じてしまうことが多い。また、日本人は謙遜することや謙虚であることなど控えめであることを美德としている。日本人は他人に褒められる、自分を褒めることが少ないため、自己肯定感を高めにくく、ストレスを溜めやすい状況がある。そのため、褒められる機会を増やすことは重要と考えられる。

そこで本研究では日常生活に関する幅広い内容のSNSの投稿に対して褒められる箇所を自動的に判別し、ユーザが嬉しいと感じる褒めリプライを自動生成するシステムを提案する。先行研究[1]においては、褒めて欲しい1単語をユーザが指定して入力する形となっていたことや、褒める対象が名詞に限られていた。本研究では、褒める対象を自動的に探すことに加えて、褒める対象を動詞にまで拡張する。

2 関連研究

2.1 SNS コメントの自動返信に関する研究

マイクロブログサービスにおいてユーザの投稿に対して自動リプライを生成する研究[2]がある。ツイート中に含まれる単語の中から特徴がある単語を抽出し、その単語を元に対話データから返信を作成する。

この研究はツイートに含まれる単語に注目し、返信を行う点で本研究と共通している。しかし、この研究では過去に投稿されたツイートを検索によって抽出し、それに対応するリプライを提案するため過去のツイートと同じものを出力する。本研究ではコメントごとに既存のものではないバリエーションのある返信を作成する。

2.2 褒め言葉を生成する研究

囲碁の初心者の方の動機づけのためにユーザの打った手を褒める研究[3]がある。この研究では囲碁AIを用いてユーザの打った手を評価し、状況に応じて褒め台詞を出力する。結果として、悪手のみを指摘する場合、動機づけを減少させる傾向にあったが、褒めることで動機づけの向上につながった。ヒューマノイド型ロボットがユーザを褒めることで学習を支援する研究[4]がある。この研究ではテスト結果に応じて動作と褒め台詞を出力し、学習者をねぎらう。結果として、学習の継続性、学習の達成感、褒められていると感じるという点に褒めることは影響を及ぼした。

どちらの研究も褒めることで能力の獲得支援を行っている。囲碁の初心者の方の動機づけのためにユーザの打っ

*連絡先： 滋賀県立大学工学部 電子システム工学科
藤原 弘貴
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: on23kfujiwara@ec.usp.ac.jp

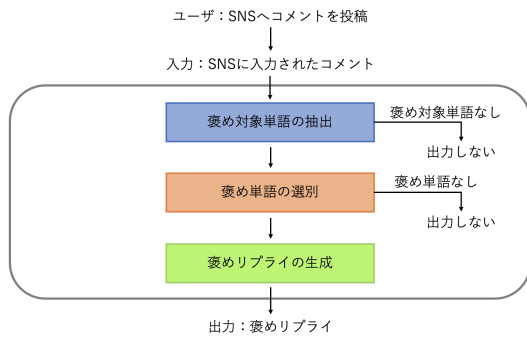


図 1: 褒めリプライの自動生成システムの構成図

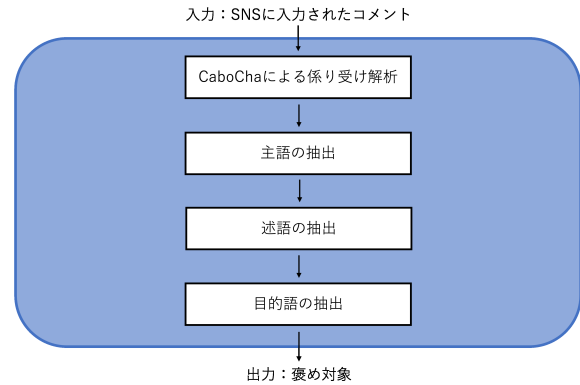


図 2: 褒め対象単語の抽出の流れ

た手を褒める研究では AI と同じ最善手を打った場合や、対局後の悪手の少なさを褒めている。ロボットがユーザを褒めることで学習を支援する研究ではテスト結果が低い点数であれば褒め言葉とねぎらいの言葉を、高い点数であればユーザの頑張りを大きな動作で褒めている。これらの研究ではシステムに褒めてもらえる場面に限られる。そこで本研究では日常生活の出来事を褒めることで幅広い場面でユーザを褒める。

3 日常の SNS コメントに対する褒めリプライの自動生成システム

3.1 システムの構成

システムの構成図を図 1 に示す。このシステムでは、ユーザが投稿した日常の SNS コメントを入力とする。入力されたコメントから褒め対象となる単語を抽出し、抽出した褒め対象を褒めることのできる単語を選別する。SNS に投稿されているコメントは、他者に見られることを前提として発信されている。そのためコメント中の動作を表す単語は他者に伝えたい自分が行ったことである。よって本研究では動作と、動作を詳しく説明する動作の対象を褒め対象とする。褒め対象単語と褒め単語から褒めリプライを生成し、出力とする。

3.2 褒め対象単語の抽出

褒め対象単語の抽出の流れを図 2 に示す。入力されたコメントに係り受け解析器 CaboCha で解析する。係り受け解析の結果より、入力されたコメントから褒め対象となる単語を抽出する。

3.2.1 CaboCha について

CaboCha とは、Support Vector Machine に基づく、係り受け解析器である。文節がその直後の文節に係るか係らないかという点に注目して解析を行う [5]。また、CaboCha における係り受け解析の手順を以下に示す。

1. 入力された文の文節全てに係り受け未定を表すタグ「O」を付与する。
2. 文末の文節を除いた「O」が付与されている文節に対して直後の文節に係るかを解析する。係る場合は「O」に係り受け確定を表すタグ「D」に置き換える。このとき、後ろから 2 番目の文節は無条件で「D」に置き換える。
3. 「O」が付与されている文節の直後の文節に「D」が付与されている場合、「D」が付与されている文節を削除する。
4. 残った文節が文末の文節のみになるまで 2, 3 の処理を繰り返す。

3.2.2 CaboCha による係り受け解析

本システムでは CaboCha を用いて、入力文の主語句、目的語句、述語句を取り出し、その結果から褒め対象を決定していく。係り受け解析の結果から主語句、目的語句、述語句を以下のように定義する。

- 係り受け解析の結果、係る先がないと出力される文末の句を述語句とする。また、述語句の中で助詞を除いた形態素解析で「動詞、自立」または「名詞、サ変接続」と出力される単語を述語とする。

- 係り受け解析の結果、係る先が述語句となる句のうち形態素解析で「係助詞」と出力される助詞を含む句を主語句とする。また、主語句の中で助詞を除いた部分を主語とする。
- 係り受け解析の結果、係る先が述語句となる句のうち形態素解析で「格助詞」と出力される助詞を含む句を目的語句とする。また、目的語句の中で助詞を除いた単語部分のうち、形態素解析で「名詞」と出力される単語を目的語とする。

3.2.3 主語の抽出

入力された文の主語を抽出する。本システムは褒めてもらうことで満足し、自己肯定感を高めることを目的としている。そのため抽出した主語が対話相手でない場合はシステムの対象外とし、「褒められません」を出力する。また日本語の文は主語が省略される場面が多くある。その理由としては直前の文と主語が同じ場合や相手と同じ場にいることが挙げられる。システムに文を入力する際はシステムと使用者が同じ場にいるため、省略されている主語は使用者となる。よって、文中に主語が存在しない場合は一人称の主語があることとして扱う。

3.2.4 述語の抽出

入力された文の述語句に注目する。述語句は主語の動作、作用、性質、状態などを動詞、名詞、形容詞、形容動詞、を用いて表される。本システムでは動作を褒めることを目的とするため述語句の中に含まれる単語のうち、形態素解析で「動詞、自立」または「名詞、サ変接続」となる動作を表す単語である述語を褒め対象として抽出する。述語句の中に動作を表す単語が存在しない場合、褒め対象単語なしとする。

3.2.5 目的語の抽出

入力された文の目的語句に注目する。目的語句は述語句に示される動作、作用、性質、状態などの対象、目的を示す。褒め対象となる動作をより詳しくするものとなるため、目的語を入力文の動作と共に褒め対象として抽出する。また日本語の語順はある程度は自由となっているが、より強調したい内容がより文末に位置する。これより動作と共に褒め対象とする目的語はより文末に近い目的語句に含まれる目的語1つとする。目的語句が存在しない、目的語句の中に目的語となる名詞が存在しない場合は目的語なしとする。

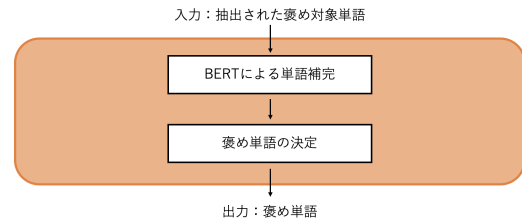


図 3: 褒め単語の選別の流れ

3.3 褒め単語の選別

褒め単語の選別の流れを図3に示す。本システムは、BERTのMLM(Masked Language Model)を用いて単語補完を行うことで褒め台詞を生成している。

3.3.1 BERTについて

本システムで用いられているBERTに関して述べる。モデルは東北大学が公開している学習済みのモデルを使用した[6]。

BERT(Bidirectional Encoder Representations from Transformers)とは自然言語処理のモデルであり、「Transformerによる双方向のエンコード表現」を意味する。Transformerと呼ばれる言語処理に秀でた深層学習モデルを用いて、文章を文頭と文末の双方向から学習する。双方向学習のためにMLM(Masked Language Model)とNSP(Next Sentence Prediction)という2つの手法を同時に行っている[7]。

3.3.2 BERTによる単語補完

本システムではBERTのMLM(Masked Language Model)を用いて褒め単語の出力を行う。MLMでは穴埋めクイズを解くように学習を行う。入力文の15%のトークン(文章の最小単位となる文字や文字列)をランダムに別のトークンで置き換えし、前後の文脈から置き換える前のトークンを予測させる。選択された15%のトークンのうち、80%は[MASK]トークンに置き換え、10%をランダムな別のトークンに変換、残りの10%は変換せず元のままとする。

本システムでは、あらかじめ[MASK]トークンを含むBERTに入力するためのテンプレートを用意する。入力用テンプレートに褒め対象となる単語を当てはめ、BERTに入力することで[MASK]トークン部分を予測し、褒め言葉を補完する。

表 1: BERT への入力用テンプレート

	目的語なし
自立動詞	あなたはとても@@に〇〇
サ変接続名詞	あなたはとても@@に〇〇する

	目的語あり
自立動詞	あなたは**△とても@@に〇〇
サ変接続名詞	あなたは**△とても@@に〇〇する

3.3.3 BERT への入力用テンプレート

表 1 に動作を表す単語の品詞, 目的語の有無ごとの BERT への入力用テンプレートを示す. ただし, テンプレート内の〇〇は動作を表す単語, **は目的語, △は目的語と同じ句の助詞とする. @@の部分 [MASK] トークンに置き換え, BERT を用いて補完し, 褒め単語を出力する.

入力用テンプレートの基本形は名詞 1 単語を褒める先行研究 [1] で使用していた「あなたの〇〇はとても@@@です」というテンプレートを参考に, 動作を褒める点に注目して「とても@@に〇〇(する)」とした.

3.3.4 褒め単語の決定

褒め単語を決定するアルゴリズムは先行研究 [1] で用いられているものと同一の以下のものを用いる.

1. BERT に [MASK] トークンが含まれたテンプレート文を入力し, [MASK] トークン部分に補完されやすい予測単語のリストを得る.
2. 予測単語の上位 20 位以内に汎用的でない褒め単語が含まれるかを調べ, 汎用的でない褒め単語が含まれている場合, その褒め単語を出力の褒め単語とする.
3. 汎用的でない褒め単語が含まれていなければ, 汎用的な褒め単語が含まれるかを調べ, 汎用的な褒め単語が含まれている場合, その褒め単語を出力の褒め単語とする.
4. 2.3. の手順において上位 20 位以内に褒め単語が存在しなければ, 順位の閾値を下げて, 21 位から 30 位の単語に関して 2.3. の手順を繰り返す.
5. 4. の手順を閾値を 50 位まで下げながら繰り返す. 50 位以内に褒め単語が見つからなければ入力単語は褒められる単語でないとし, 褒め単語は出力しない.

また, 出力として [UNK] が出力されることがある. これが出力された際は適切に補完できる単語がそれ以

表 2: 褒めリプライのテンプレート

	目的語なし
自立動詞	あなたは@@に〇〇ね
サ変接続名詞	あなたは@@に〇〇するね

	目的語あり
自立動詞	あなたは**△@@に〇〇ね
サ変接続名詞	あなたは**△@@に〇〇するね

降にはないとし, それ以降の探索は行わない. 加えて, 全く同じ予測の出力をする単語が複数ある. 学習時にトークン分けされた際に, 1 単語として学習されていないであろう単語などを入力した時も同様の出力をする. それらの単語は学習がされていないなどの理由で適切に単語を補完されていないとし, 褒め単語を出力しない.

3.4 褒めリプライの生成

出力用テンプレートは動作を表す単語の品詞, 目的語の有無ごとに設定した. 表 2 にそれぞれの場合の出力用テンプレートを示す. 褒めることで相手を喜ばせ, 自己肯定感を上げることを目的とするため語りかけるような口調とした. また単語が補完されない場合には「褒められません」, 入力された文に褒め対象がない場合は「褒め対象がありません」を出力することで入力された文が褒められない原因を示す出力を行う.

例えば, 「今日の授業が分からなかったので数学を勉強した」という SNS に投稿されたコメントを入力した際には, 係り受け解析の結果, 主語が存在しないため, 一人称の主語がある場合として扱う. また, 述語は「勉強」, 目的語は「数学」となる. 述語, 目的語が BERT への入力テンプレートに当てはめられて「あなたは数学をとても@@に勉強する」という文において, @@に当てはまる単語を BERT の MLM を用いて補完する. ここで補完された単語が「熱心」となるため, この単語を褒めリプライの出力用テンプレートに当てはめた「あなたは数学を熱心に勉強するね」を褒めリプライとして出力する.

4 日常の SNS コメントに対する褒めリプライの自動生成システムの評価実験

本章では, 実装した日常の SNS コメントに対する褒めリプライの自動生成システムが出力した褒めリプライにより, ユーザに嬉しいと感じてもらえるかを検証するために行った実験について述べる.

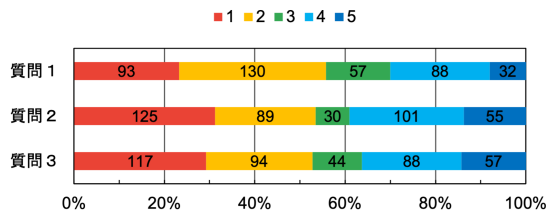


図 4: 褒めリプライに関する質問の回答結果

本実験により、システムの出力したリプライに対して、ユーザが嬉しいと感じる褒めリプライの生成ができるか、入力されたコメントの褒めて欲しい部分を褒めているか、得られた褒めリプライが褒めリプライとして成り立っているかを確認する。

4.1 実験方法

理系の大学生、大学院生 10 名を被験者として、「SNS の日常生活に関するコメントとシステムの生成した褒めリプライの対話文」を、全部で 40 セット用意して、以下の手順で実験を実施した。

1. SNS から収集した日常生活に関するコメントとシステムが出力した褒めリプライの対話文を読む。
2. 質問 1: 「システムの出力した褒めリプライを、コメント投稿者がどれくらい嬉しく感じますか？」に回答する。(「1. とても嬉しい」から「5. 全く嬉しくない」の 5 段階評価)
3. 質問 2: 「システムの出力した褒めリプライが、コメント投稿者が褒めて欲しいところを褒めようとしていますか？」に回答する。(「1. 褒めて欲しいところを褒めようとしている」から「5. 褒めて欲しいところとは違うところを褒めようとしている」の 5 段階評価)
4. 質問 3: 「システムの出力した褒めリプライのみを見た場合、褒め言葉として自然に感じますか？」に回答する。(「1. 自然な日本語だと思う」から「5. 日本語としては不自然だと思う」の 5 段階評価)

4.2 結果と考察

図 4 に褒めリプライに関する質問の回答結果を示す。質問 1、質問 2、質問 3 の全ての質問において 50% 以上は、1 または 2 と回答された。また、いずれの質問も 3 という回答が少ないが 4 という回答の多さが目立った。このことから、一定のコメントに対しては本システムの有用性が示されるとともに、入力コメントは本シ

テムが対応できるコメントと対応できないコメントに分かれていることが考えられる。

システムの出力した褒めリプライを嬉しく感じるかという質問 1 に関する考察を述べる。質問 1、質問 2、質問 3 の回答の値に大きなばらつきはほぼなく、質問 2、質問 3 の評価に応じて質問 1 の評価が決定される考えられる。中でも質問 2 のシステムの出力した褒めリプライが、コメント投稿者が褒めて欲しいところを褒めようとしているかの評価が質問 1 の評価に大きく影響すると考えられる。褒めて欲しいところとは違うところを褒めようとしている褒めリプライが日本語として不自然になり、その結果システムの出力した褒めリプライを嬉しく感じないことに繋がる。そのため、システムの出力した褒めリプライを嬉しく感じてもらうためにはコメント中の褒めて欲しい部分を自然な褒め言葉で褒めることが必要である。

システムの出力した褒めリプライが、コメント投稿者が褒めて欲しいところを褒めようとしているかという質問 2 に関する考察を述べる。褒めて欲しいところを褒めることができていると評価されたコメントに注目する。他に褒めるべき対象があることが褒めて欲しいところを褒めることができている理由となる。

他に褒めるべき対象となる単語として、コメントの中で動作を修飾する(規模、頻度など)単語がある。例として「10 キロ走った」の「10 キロ」や「たくさん作った」の「たくさん」のように動作をどの程度行ったかを表す単語がある。これらの単語は「動作を修飾する単語」+「述語」として褒め対象の候補になると考えられる。動作を修飾する単語を抽出するためには述語句にかかる目的語句以外の文節も考慮する必要がある。述語句に一番近い句に動作を修飾する単語が存在する場合は優先して褒め対象とするアルゴリズムへの改良や、動作を修飾する単語が褒め対象として抽出された場合のテンプレートを用意する必要がある。

また、コメント中の述語ではなく、1 名詞が褒めるべき対象となる場合がある。述語のみを褒め対象とするのではなく、述語と 1 名詞についてそれぞれ BERT から褒め単語を生成し、より補完される可能性の高い褒め単語を出力した方を褒め対象することで褒めるべき対象となっている名詞を適切に褒めることができると考えられる。今回の評価実験でこれらの改善方法を用いると質問 2 で評価点の平均が 3 以上となっている 16 個の対話文のうち 9 個の対話文の評価を改善することができる。動作の規模や頻度を表す単語と共に述語を褒め対象とする場合は 7 個、コメント中の名詞を褒め対象とする場合は 2 個の対話文に対応できる。

システムの出力した褒めリプライのみを見た場合、褒め言葉として自然に感じるかという質問 3 に関する考察を述べる。褒め言葉として自然に感じない理由として出力されたリプライが日本語として不自然だというこ

とがある。この原因として BERT への入力テンプレートや褒めリプライの出力のテンプレートが適切でないことや褒め言葉自体が適切でないことが考えられる。

BERT への入力テンプレートや褒めリプライの出力のテンプレートとして「@@に～する」の他に「～するなんて@@」というテンプレートが考えられる。褒め対象となる述語によっては既存テンプレートよりも「～するなんて@@」のテンプレートの方が褒め単語となる単語が容易に考えられるものがある。そのため、述語ごとに適切なテンプレートを使い分ける、複数のテンプレートを用いて共通する褒め単語を出力する、複数のテンプレートを用いて補完されやすい褒め単語を出力するといったアルゴリズムへの改良が必要である。

また褒め言葉自体が適切でない理由としては、先行研究で用いられた褒め単語をそのまま利用していることが挙げられる。先行研究では、名詞 1 単語を褒め対象とした場合の褒め言葉を用意しているため、本研究における述語用の動詞を褒め対象とした場合の褒め単語としては適切でないものがあつたと考えられるため、述語用の褒め単語の再検討が必要と考えている。

5 おわりに

SNS 投稿したコメントを褒めることで自己肯定感を高めてもらうことを目的として日常の SNS コメントに対する褒めリプライの自動生成システムを実装した。また、日常の SNS コメントに対する褒めリプライの自動生成システムがユーザの自己肯定感を高めることに有用であるかを確認するために評価実験を行った。

この実験において、一定のコメントに対しての本システムの有用性が示された。また、本システムが対応できるコメントと対応できないコメントが存在することが確認された。

今後の課題としては、コメント中の適切な褒め単語を抽出するアルゴリズムへの改良や、対応するテンプレートの追加が求められる。また、褒め単語を生成するための BERT への入力テンプレートの追加や動詞用の褒め単語の検討をすることでより適切な褒め単語を出力させる必要がある。

褒めリプライの生成が日常の SNS コメントだけでなく、教育現場や職場など褒められることで自身のモチベーションに繋がる場面で応用されることで、快適な環境構築の足がかりになることを期待したい。

参考文献

- [1] 堀紘之輔, 砂山渡: BERT を用いた単語補完による褒め台詞生成システム, 第 28 回人工知能学会イ

ンタラクティブ情報アクセスと可視化マイニング研究会, SIG-AM-28-04, pp.20 – 25 (2022)

- [2] 森下雄太, 渋谷翔太, 佐藤隆士: マイクロブログサービスにおける自動リプライシステムの考案, 電子情報通信学会, 第 15 回情報科学技術フォーラム, D-027 (2016)
- [3] 若林広志, 伊藤毅志: 囲碁初心者の動機づけを目的とした着手を褒めるシステム, 情報処理学会研究報告, Vol.2020-GI-43, No.2., pp.1 – 8 (2020)
- [4] 平野愛里, 松田晃一: ヒューマノイド型ロボットを用いた褒める行為に着目した学習支援システムの試作と評価, 情報処理学会研究報告, Vol.2019-CE-148, No.15, pp.1 – 8 (2019)
- [5] 工藤拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.43, No.6, pp.1834 – 1842 (2002)
- [6] 東北大学乾研究室: Pretrained Japanese BERT models released / 日本語 BERT モデル公開: (URL) <https://github.com/cl-tohoku/bert-japanese> (2023 年 1 月 24 日確認)
- [7] Jacob Declin Ming-Wei Chang Kenton Lee Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, In proceedings of the 2019 Conference of North America Chapter of Association for Computational Linguistics: Human Language Technologies, Vol.2 (2019)

災害情報収集における SNS のメディア特性に関する一検討

Media Characteristics of Social Network Services in Collecting Disaster Information

安尾 萌 *

Megumi Yasuo

森野 穰

Yutaka Morino

松下 光範

Mitsunori Matsushita

関西大学大学院総合情報学研究科

Graduate School of Informatics, Kansai University

Abstract: 災害発生時、被害報告や救援要請は復興支援を行う上で欠かせない情報である。災害情報は平常時よりも情報の鮮度がより重視されることから、速報性に優れた情報インフラとして SNS を活用し、災害復興支援に利用するための研究が存在する。従来研究では、災害情報の獲得に利用する情報ソースが偏っている現状がある。この状況は SNS の構造や利用者数などが原因として考えられるが、災害に関する情報インフラが単一のサービスに依拠する状況は必ずしも好ましいものではない。本稿では災害情報収集に適したメディア特性とそのあり方について、現行の SNS のメディア特性、および災害発生時に必要な情報に関する要件の整理を通じて論考する。

1 はじめに

我が国は、台風や地震といった自然災害が発生しやすい地理的特徴を有しており、しばしば災害大国と称される。2022 年においても、11 月に発生した台風 14 号、15 号による風水害や 12 月に発生した雪害など、複数の大規模自然災害に見舞われたことが報告されている^{*1}。現代において災害が発生した際、災害に関する情報を発信するのはマスメディアや政府広報に限らない。Web の発展と普及に伴って、被災者やその周囲の人々といった個人が、災害に関する情報を発信、共有するようになった。これらの情報は今日、災害対策を目的としたソーシャルメディアの情報の活用を試みるいくつかの研究において注目されている [1, 3]。

災害発生時におけるソーシャルメディア情報の活用を目的とした従来研究の多くは、マイクロブログサービスである「Twitter」をデータソースとして利用している。Twitter は現在、最も広く普及した SNS の一つである。内閣府の SNS 活用における調査 [29] では、若年層が利用者の中心となっているだけでなく、40～50 代の中年層においても 30% 前後が利用していることが明らかに

なっており、日本における Twitter は社会的な情報インフラになりつつある。また Twitter は運営会社が公式に提供している API^{*2}が存在し、投稿されたデータを容易に獲得することが可能である。そのため、多数の学術分野における重要な研究資源としても広く利用されている [18]。災害情報収集においても、Twitter を情報収集源とする研究は多い [5, 8]。

しかし、Twitter は企業が提供している SNS サービスであり、運営する企業の意向次第で環境が大きく変化する懸念がある。実際、2023 年 2 月に公表された API の有料化に関するニュース^{*3}を受け、Twitter と連携する様々なサービスの停止が懸念されるなど、大きな影響が見られた [31]。災害情報についても、例えば約 318 万人のフォロワーに地震情報を発信してきた Twitter アカウント「地震速報」 (@earthquake_jp) が API 有料化の影響により閉鎖を検討していることが報じられている [35]。加えて、こうした影響はユーザの Twitter 離れ [30] を加速させる恐れがあり、ひいては情報収集ソースとしての価値の低減に繋がることが懸念される。こうした懸念を考慮すると、災害に関する web 上の情報インフラを Twitter に依存する状況は必ずしも好ましいとは言えない。

* 連絡先：関西大学総合情報学部

〒569-1095 大阪府高槻市霊仙寺町 2-1-1

E-mail: k290993@kansai-u.ac.jp

^{*1}内閣府：災害情報、防災情報のページ, <https://www.bousai.go.jp/updates/index.html> (2023/3/1 確認)

^{*2}Twitter API: <https://developer.twitter.com/en/docs/twitter-api> (2023/3/1 確認)

^{*3}@TwitterDev: <https://twitter.com/TwitterDev/status/1621026986784337922> (2023/3/1 確認)

そこで本稿では、web サービスに投稿される情報の性質を災害情報収集という観点から整理し、現在のTwitter が災害情報収集において優れている点を踏まえ、他メディアの利活用及び課題解決のための方策を検討する。

2 ユーザ投稿情報

ソーシャルメディアをはじめとしたユーザ投稿型のweb サービスでは、災害発生の如何に関わらず毎日大量の情報が投稿される。本研究ではweb サービスの利用者によって投稿された情報群のことを「ユーザ投稿情報」と定義している。ユーザ投稿情報は「ゲームに関する情報」や「芸能情報」といったトピックを基準に扱うweb サービスと、「動画」や「音楽」といったモダリティを基準に扱うweb サービスに大別できる。トピックを基準にしたweb サービスは、各専門のトピックについて興味のある人々が訪れ、各トピックから逸脱しないという制約のもとで交流する形式が一般的である。代表的なものとしては、製品の口コミ投稿サイトやレシピの共有サイト、ゲームに関する情報を専門に扱う掲示板などが挙げられる。モダリティを基準にしたweb サービスは、テキストや画像、動画、音声といった情報の形態を基準としており、トピックに関する制約は存在しない。代表的なものとしては、動画を中心としたYoutube や画像を対象にしたInstagram などが挙げられる。これらのweb サービスの性質が投稿者、情報収集者にとってどのような影響があるかについて整理する。

トピックを基準にした投稿型web サービスは、例えばゲーム、化粧品、地元の情報など、話題が固定されていることを前提とした設計が採用されている。この設計は、特定のトピックについて会話をする、あるいは特定のコンテンツを評価するといった目的において利用しやすい形式のweb サービスである。例えばゲームのレビューを専門的に投稿するweb サービスにおいては、投稿者が遊んだゲームを軸とした話題が投稿されることになり、同一の目的を持つ投稿者同士が集まりやすい。そのため、トピックを基準にしたweb サービスは、企画するトピックが明確である利用者同士の投稿を中心に構成されているといえる。こうしたweb サービスのデータは、専門性の高い語彙の獲得や知識の構築に利用される[6, 15]。災害というトピックを基準にした投稿型web サービスとしては、携帯電話のキャリア会社や一部の自治体が提供している災害用伝言板や災害用掲示板サービス、一部の災害アーカイブサービスが挙げられる。例え

ば「災害アーカイブぎふ」*4 は投稿型web サービスの形式を採用している災害アーカイブサービスである。このweb サービスには主に位置情報付きの写真画像が利用者から投稿される。蓄積された災害情報を活用したオンラインワークショップなども実施されている[22]。

モダリティを基準にしたweb サービスは、一般的にコンテンツの形態のみを規定するものであり、内容に関する規定がない場合がほとんどである。そのため投稿されるトピックの自由度が高く、同一のweb サービス内であっても投稿される内容に統一性がない。そのため、これらのweb サービスを対象とした研究は、モダリティそのものの性質を対象とした研究が中心となる。特定のトピックを対象とした研究のデータソースとして使用する場合は、タグやキーワードによって対象となるデータを絞り込んだ上で利用される。ただし近年では、動画と動画の発話の書き起こしデータを併用する、投稿された画像とそのキャプションを紐づけるなど、モダリティ間の紐付けを行うことで、従来では実現が難しかった試みが可能になりつつある[12, 25]。

情報収集をするという観点で各web サービスを外観すると、目的とするトピックが明確な場合はトピックを基準にしたSNS へアクセスすることが望ましい。しかし、トピックを基準にしたweb サービスは情報量が限られ、更新速度も芳しくないことがしばしばある。これは、トピックを基準にしたSNS の利用シーンや投稿内容への要求が影響していると考えられる。トピック基準の投稿型web サービスにおいて安定した利用者のいるサービスは、食事や化粧品といった多くの人にとって馴染み深いトピックをカバーしているケースが多く、このトピックが専門的になるほど投稿の難易度が上がる。また、Wikipedia のように閲覧者が多く情報の需要が高いweb サービスにおいては、投稿内容に一定の品質が求められるケースもしばしばある。以上の点から、トピック基準のweb サービスはモダリティ基準のweb サービスよりも投稿内容についての要求水準が高い。特に、災害発生時のような突発的なイベントの場合、トピックを基準にしたweb サービスのプラットフォームはそもそも成立しづらく、周知されにくい問題もある。地震や台風など、繰り返し起きる災害については、災害発生前にweb サービスを用意することは可能であるものの、平常時にこうしたweb サービスを日常的に活用することは考えにくく、サービス利用者数は限定的になり、災害発生時の主要な情報発信・共有の場になることは考えにくい。そのため、災害に関する情報はトピックの自由度が

*4 <http://gifu.shinrokuden.irides.tohoku.ac.jp/ecom/index.php>(2023/3/2 確認)

高いモダリティ基準の web サービスに投稿されやすいと考えられる。

3 災害情報と SNS 特性

本章では、前章におけるソーシャルメディアの分類から Twitter というプラットフォームの利点と欠点を整理し、災害発生時に必要な情報とその要件について各段階別に考察する。

3.1 投稿型 web サービスとしての Twitter の特性

投稿内容に一定の品質を要求するという特徴から、トピック基準の web サービスはモダリティ基準の web サービスよりも投稿内容に関する編纂コストを投稿者に負担させているといえる。例えばゲームについて興味のある投稿者が、ゲームについて記述したテキストを何らかの投稿型 web サービスに投稿する場合を考える。テキストモダリティを基準にした web サービスであれば、表現媒体がテキストである以上はそのまま投稿することができる。しかし、ゲームというトピックを基準にした web サービスに投稿する場合は、無関係な記述の削除など、記述内容の推敲をする必要がある。情報収集者にとっては、この特性によって一定の品質が保障された情報を一つの web サービス上で収集することができるというメリットがある。一方で情報の発信者にとっては、発信内容の品質向上のために時間的コストを要求される。このコストが要求されず、投稿者にとって負担が少ないという点は、モダリティ基準の web サービスの魅力の一つであるといえる。この特徴は他の投稿型 web サービス、とりわけ SNS に多く見られる。現在一般的に利用されている SNS は、「ソーシャルメディアのうち人間関係の構築と維持、および情報共有を目的としたサービス」と言える [33]。一般的な SNS に標準的に付与されている「いいね」機能やシェア機能も、コミュニケーションを促進させてアカウント同士の繋がりを構築する手がかりとして利用される。そのため、SNS のプラットフォームは投稿内容の品質の担保よりも、投稿者が気軽に投稿できることを志向した仕組みが重視されている。例えば Twitter の基本機能である文字数制限をはじめ、Instagram における「ストーリー」や Twitter における「フリート」のように、一定時間で投稿が自動削除される機能が挙げられる。

以上を踏まえると、発信者が Twitter を災害時に利用することの利点は、Twitter が対話性を持つ SNS であるという点にある。上述のように、SNS は構築された人

間関係をベースにした web サービスである。これは自身の投稿が馴染みのあるアカウントから徐々に広く伝播していく形態であり、情報共有と同時に仲間との対話を行う側面が強い。Twitter や Instagram などの代表的な SNS を対象とした研究では、メンタルヘルスへの影響分析に関する研究が盛んに試みられているほか [16]、SNS を介したコミュニケーションによるメンタルケアの試みも存在する [4]。この点を考慮すると、SNS に投稿される災害情報は通報や救援要請よりも、仲間内とのコミュニケーションを通じた安否の報告や精神的安定を企図して発信されている情報であると認識の方が適切である。

3.2 Twitter 利用における諸課題

情報収集の観点からこの課題を観察すると、Twitter を利用することは必ずしも適切ではなく、課題解決において最も信頼性の高い情報へ第一にアクセスすることが最も望ましい。前述したように、Twitter に流れる情報はコミュニケーションを前提としている。そのため書き込まれる情報の中には誇張や虚偽の情報、誤情報がしばしば含まれることがあり、これらの情報を検出することは大きな課題となっている [17]。例えば、Mendoza らは 2010 年に発生したチリでの大地震における Twitter の影響を調査し、地震直後の公式な情報源からの情報が乏しいときに、Twitter 上に投稿され拡散されたいくつかの噂が、現地住民の混乱や不安感を増大させたことを報告している [9]。重要な判断をするために正確な情報を収集する必要がある場合は、政府広報やマスメディアからの放送のように内容を精査、集約された情報群や、災害伝言板のように発信元が明確な情報群を対象とすることが合理的だろう。災害時における SNS からの情報収集では、専属の職員を配備する必要がある、人的負担が大きいという問題が指摘されている [13, 32]。これは、デマや事実と異なる投稿が存在することで情報の信頼性の担保が難しい [34] ことに加え、被災地以外からの無関係な投稿による情報過多が発生し、人目による情報確認を行うことが求められるためである。

この情報過多の原因の一つに、SNS の日常的な利用の多様性が挙げられる。SNS 上では様々な情報のやり取りが日常的に行われている。特にエンタテインメントと SNS の親和性は高く、通常時における Twitter 投稿の 6 割はエンタテインメント情報であると言われている [19]。実際、災害時においても趣味や娯楽に関する情報が多く投稿されており、自治体による情報収集の際に、迅速な対応を困難にしている。SNS からの災害情報の収集では、災害に関するキーワードやハッシュタグ

(例えば「救助」「被害」「支援」など)が用いられるが、それらはゲームやアニメなどでも頻繁に用いられる語彙であるため、選択的な情報収集を難しくしている。災害時の救助や支援を求めるという行動は“ユーザが要求している”という点でゲーム内行動の構造と類似しており、自治体が行っているようなキーワードマッチングでの情報収集では分離することが困難である。

森野らはこうした状況に対処するため、令和2年7月3日から31日にかけて発生した「令和2年7月豪雨」を分析対象とし、どのようなコンテンツが混淆しやすいのかを分析するとともに、それらのコンテンツが分離可能かについて検討している [36]。その結果、ゲームに関する情報は分離できる可能性が高いものの、グッズに関する情報は難しいことを明らかにした。

以上のように、Twitter 上の情報の利用にはいくつかの課題が存在し、獲得したデータを「災害情報」として汎用的に利用することは現段階では難しい。Twitter 上の情報を利用した災害情報収集について考察するには、災害情報にどのような区分が存在し、どのようなシーンにおいて Twitter の情報を活用することが適しているかを整理する必要がある。

3.3 災害発生時に必要な情報とその要件

災害発生時から復旧に至るまでに必要な情報は、災害発生時からの各段階によって変化する。消防庁は、大規模災害発生時の災害対策について「警戒段階」「発災段階」「被災者支援段階」の3段階に分類しており、各段階別に収集すべき情報をまとめている [24]。警戒段階は災害発生の前段階に位置し、避難勧告や避難指示の発令、災害対策本部の設置が行われる。この際市町村では、天候の情報や河川の水位、住民の避難状況などについて情報収集を行う。これらの情報源は主に各観測所が中心となる。発災段階において最も優先される情報は、人命救助に関わる情報である。特に災害発生時から72時間は人命救助の成功の可否に関わる重要な時間であり [38]、この間に必要となる情報は「どこに負傷者がいるか」また「いかにして救命するか」に関する情報が最優先になる。被災者支援段階では、災害による具体的な被害状況に関する情報が優先されるようになる [14]。これは避難所の運営や管理に関わる情報や住宅の損壊などの情報といった被災者の生活に関わる情報だけでなく、通行が困難になった道路や水道の損害など、社会的なインフラに関わる情報も含まれる。これらの情報の集約は市民からの通報のほか、市の職員による巡回などによって行われ、地域によっては対象エリアが広範囲にわたり、多くの人的、時間的リソースを必要とすることも多い。以上の点

を踏まえると、各段階において必要な情報の要件が大きく異なることがわかる。発災段階では直接人命に関わることや、災害発生時という状況下で利用可能なリソースが限られることを考慮すると、この段階で収集する情報は救助を要請する人物が明確であり、一般的な SNS のプラットフォームから区別可能であることが求められる。被災者支援段階では、広域を対象とした情報を大規模に収集する必要がある。そのため、多人数で協調的に情報を集約する仕組みが有効であるといえる。

4 ユーザ投稿情報の活用事例

LINE および Facebook では、クローズドな空間に投稿された情報の活用がされている。自治体による LINE 利用の実例として、自治体の公式アカウントによる被災者への情報発信や、LINE の基本的なメッセージ機能を用いたユーザ対チャットボットによる双方向のコミュニケーションが挙げられる。令和元年台風15号の際、千葉県は実際に AI チャットボットを活用した情報発信を行い、被災者からの問い合わせに対応する形で情報提供を行っていた*5。また、Facebook は、独自に災害時情報センターを設けている*6。この災害時情報センターでは、被災したユーザは自身の安否を投稿することができ、災害の影響を受けた地域にいる友達の安否が確認可能である他、サポートが可能なユーザは実際にサポートする意思を表明することができる。ユーザ同士の双方向コミュニケーションを促すことにより、ユーザ同士の助け合いを促している。

総務省所管の国立研究開発法人情報通信研究機構(NICT)は災害情報収集システム「DISAANA」を開発している [11, 20]。DISAANA は Twitter の投稿をリアルタイムで分析することで、どこで何が起きているのかを抽出するシステムである。5W1H を含む問かけ文によるキーワード検索機能や、地点をクエリとした情報収集機能を提供している。SNS に散見されるデマや、事実関係が不明瞭な情報には、その情報に対して矛盾する投稿を提示することで、システム利用者自身に真偽判断を促している。また、Castillo らは、ユーザの投稿・リツイート行為や投稿内容、参照する外部リソースを特徴量として投稿された一連の Tweet の信頼性を判定する技術を提案し、70-80% の精度で判定できることを示している [2]。

宋らは、平成30年7月豪雨の際、NHK 記者により確

*5 現在はサービス運用を終了している。

*6 Meta: 災害時情報センターのご案内, <https://about.fb.com/ja/news/2014/10/safetycheck/> (2023/3/1 確認)

認された救助要請ツイートを分析することで、救助要請ツイートの特徴を (a) 番地までの詳細な住所 (b) 町名に加えて、団地名・施設名・建物名を記載した住所 (c) 関係性を表す語または「友達」、とした [27]。この特徴のうち (a) と (b) の有効性を検証した結果、この特徴を含むツイートに救助要請ツイートが含まれる確率は 7.94% であった [28]。これにより、救助要請ツイートを自動的に検出する可能性を示した。ただし、これら検証は、全体のデータ量が少なく、結果は限定的である。小林らは、RNN を用いた噂や不確実な情報の早期検出のために、ユーザ分類問題で有用とされている過去ツイートを活用することでユーザ特徴を表現した [21]。リツイートしたユーザ特徴を多変量時系列化することで、ツイートの伝播経路を構築した。佐藤らは、Twitter Japan 社が推奨している救助要請を行う際の記載方法の一つである「#救助」というハッシュタグが付与されたツイートの分析を行った。2017 年 7 月九州北部豪雨災害と 2018 年西日本豪雨災害時のツイートを比較し、#救助ツイートの実態を定量的に明らかにした [23]。その結果、実際に災害の状況を記載した#救助ツイートは、それぞれの災害で 7.6%, 16.5%(リツイート含む) となり、その大多数がニュース記事や被災地域外からの善意による投稿であることが確認された。

5 議論

前章までの議論をもとに Twitter を災害情報の収集に利用するという観点から、利用に適している側面、および課題について整理し、実際の活用における提案を行う。

Twitter はしばしば災害発生時に利用者によって被災報告が書き込まれる。この要因として、日常的に利用されていてユーザが利用形式に馴染んでおり、情報の編纂コストなどの制約がないことから、利用者が容易に投稿可能であること、たくさんの利用者を通じて構築されたネットワークが存在し、そのネットワークを通じたコミュニケーションや投稿から獲得できるフィードバックによって安心感を獲得できること、などが考えられる。ただし、ユーザ投稿情報を災害情報の収集に利用するにあたっては、いくつかの課題が存在する。例えば被災者以外のアカウントから発信される日常的な投稿が災害情報と混淆するコンタミネーションの問題や、重要な情報が重要度の低い情報によって埋もれてしまう問題、また正しくない情報や古い情報が共有され続ける問題などである。以上の課題を踏まえて、ユーザ投稿情報を災害対策に活用する方策を検討する。

5.1 広範囲の物的被害状況に関する情報収集

被災者支援段階における物的被害状況の収集は、対象が広範囲にわたることから被害箇所の把握に多くの時間を要することがある。こうした被害状況に関する情報は市民からの通報のほか、自治体職員によって構成される災害対策本部などが、直接被災地域を巡回して収集、管理することも多い。こうしたタスクに対しては、多数の利用者が存在する SNS を活用することで、効率的な情報収集を可能にすることが期待できる。この仕組みの一例として、位置情報と写真を紐づけた災害被害箇所収集ツールが存在する [37]。これは写真と位置情報を紐づけた簡易型のツールであり、ツールの利用者が被害箇所の発生場所と被害の内容を写真と共に記録することができる。データ登録の際に携帯端末の位置情報と撮影時の方向を取得することで、収集された情報を自動的にマッピングできる設計となっている。

5.2 ユーザ投稿情報に対する一時的な投稿内容の認証

信頼性の高い情報を獲得する一般的な方法として、情報の発信者がどこにいるのかを明らかにするというものが考えられる。災害発生時、携帯電話の通信事業者が設置する災害伝言ダイヤルや災害掲示板はその一例である。一部の web サービスにおいても、クレジットカードの連携や SMS 認証を経由することで、発信者の年齢認証や発信内容と個人の関連付けに利用するサービスが存在する。このような仕組みを利用して、既存の web サービス上で一時的に電話番号などによる緊急の投稿を行う仕組みを構築し、一般的な投稿とは区別された情報として収集することで、利用者の情報の信頼性を担保することが可能になる。

5.3 収集した情報の協調的な峻別

この仕組みは、複数人で協調的に情報の収集、峻別に取り組むための方策を構築することで、時間的、資源的制約下におけるタスク遂行の全体最適化を目指す方法である。災害情報の収集を行う際、その人員は個別にタスクを持っていることがあるが、複数のタスクに関わる情報について各人が個別にその内容を吟味することは時間的コストがかかる。この課題に対して、複数人で協調的に情報の収集、峻別を行う仕組みを構築することで、情報の集約を効率化することを目指す試みが存在する。この仕組みは「協調的情報トリアージ」と呼ばれてい

る [26]. ソーシャルメディアの情報を対象にした協調的な情報トリアージのためのツールを開発する試みも存在し [7], 膨大な量の情報をもとにした課題解決や意思決定の効率化が期待できる。

6 おわりに

本稿では, 災害発生時に必要な情報に関する要件, および現在の SNS のメディア特性の整理を通じて, 災害情報収集において必要となるメディア特性とそのあり方について検討した. 災害発生時に必要となる情報の要件から, SNS 上の情報収集を人命救助に利用するには発信者と紐づいた信頼性の担保が必要であること, 現行の SNS を利用する場合は物的被害の状況に関する情報収集に適していることを述べた. ソーシャルメディアは災害情報収集だけでなく, 災害発生時における避難行動などを促す研究 [10] も試みられている. 災害発生時において, SNS 上から災害情報を収集可能にすることは, 災害分析や今後の災害対策に関する取り組みにおいて大きな貢献になると考える。

参考文献

- [1] Alexander, D. E.: Social Media in Disaster Risk Reduction and Crisis Management, *Science and Engineering Ethics*, Vol. 20, No. 3, pp. 717–733 (2014).
- [2] Castillo, C., Mendoza, M. and Poblete, B.: Information Credibility on Twitter, *Proceedings of the 20th International Conference on World Wide Web*, p. 675–684 (2011).
- [3] Chair, S., Charrad, M. and Saoud, N. B. B.: Towards A Social Media-Based Framework for Disaster Communication, *Procedia Computer Science*, Vol. 164, pp. 271–278 (2019).
- [4] Chmielowska, M., Zisman-Ilani, Y., Saunders, R. and Pilling, S.: Social network interventions in mental healthcare: a protocol for an umbrella review, *BMJ Open*, Vol. 11, No. 12 (2021).
- [5] Crooks, A., Croitoru, A., Stefanidis, A. and Radzikowski, J.: # Earthquake: Twitter as a distributed sensor system, *Transactions in GIS*, Vol. 17, No. 1, pp. 124–147 (2013).
- [6] Guo, X., Liang, L., Liu, Y., Weng, H. and Hao, T.: The Construction of a Diabetes-Oriented Frequently Asked Question Corpus for Automated Question-Answering Services, *Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, p. 60–66 (2020).
- [7] Herceg, P. M., Allison, T. B., Belvin, R. B. and Tzoukermann, E.: Collaborative exploratory search for information filtering and large-scale information triage, *Association for Information Science and Technology*, Vol. 69, No. 3, pp. 395–409 (2018).
- [8] Ibrahim, R., Elbagoury, A., Kamel, M. S. and Karray, F.: Tools and approaches for topic detection from Twitter streams: survey, *Knowledge and Information Systems*, Vol. 54, pp. 511–539 (2018).
- [9] Mendoza, M., Poblete, B. and Castillo, C.: Twitter under Crisis: Can We Trust What We RT?, *Proceedings of the First Workshop on Social Media Analytics*, p. 71–79 (2010).
- [10] Mirbabaie, M., Ehnis, C., Stieglitz, S., Bunker, D. and Rose, T.: Digital nudging in social media disaster communication, *Information Systems Frontiers*, Vol. 23, pp. 1097–1113 (2021).
- [11] Mizuno, J., Tanaka, M., Ohtake, K., Oh, J.-H., Kloetzer, J., Hashimoto, C. and Torisawa, K.: WISDOM X, DISAANA and D-SUMM: Large-scale NLP systems for analyzing textual big data, *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 263–267 (2016).
- [12] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. and Sutskever, I.: Learning Transferable Visual Models From Natural Language Supervision, *arXiv*, <https://arxiv.org/abs/2103.00020> (2021). (2023/3/1 確認).
- [13] Sakurai, M., Majchrzak, T. A. and Latinos, V.: Towards a Framework for Cross-Sector Collaboration: Implementing a Resilience Information Portal, *Information Systems for Crisis Response and Management in Mediterranean Countries* (Dokas, I. M., Bellamine-Ben Saoud, N., Dugdale, J. and Díaz, P.(eds.)), Cham, Springer International Publishing, pp. 177–192 (2017).
- [14] Sakurai, M. and Murayama, Y.: Information technologies and disaster management — Ben-

- efits and issues —, *Progress in Disaster Science*, Vol. 2, Article 100012 (2019).
- [15] Ueda, M., Taniguchi, Y., Li, D., Siriaraya, P. and Nakajima, S.: A Research on Constructing Evaluative Expression Dictionaries for Cosmetics Based on Word2Vec, *The 23rd International Conference on Information Integration and Web Intelligence*, p. 80–86 (2022).
- [16] Valkenburg, P. M., Meier, A. and Beyens, I.: Social media use and its impact on adolescent mental health: An umbrella review of the evidence, *Current Opinion in Psychology*, Vol. 44, pp. 58–68 (2022).
- [17] Zhou, X., Zafarani, R., Shu, K. and Liu, H.: Fake News: Fundamental Theories, Detection Strategies and Challenges, *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*, p. 836–837 (2019).
- [18] 植田麦: 研究資源としての Twitter, Vol. 564, 明治大学教養論集刊行会, pp. 17–41 (2022).
- [19] NEC ビッググローブ株式会社: 東日本大震災におけるツイッターの利用状況について～新たな情報摂取・共有スタイルの定着～, <https://www.biglobe.co.jp/pressroom/release/2011/04/27-1> (2011). (2023/3/1 確認).
- [20] 大竹清敬: 災害時における DISAANA、D-SUMM の活用～DISAANA・D-SUMM と防災チャットボット SOCDA～, https://www.soumu.go.jp/main_content/000672984.pdf (2020). (2023/3/1 確認).
- [21] 小林鼓, 藤田桂英: ユーザの過去ツイートを用いた噂の早期検出, 情報処理学会研究報告, Vol. 2020-ICS-199, No. 9, pp. 1–8 (2020).
- [22] 小山真紀, 荒川宏, 伊藤三枝子, 平岡祐子, 柴山明寛, 井上透: 災害アーカイブぎふを活用したオンラインワークショップ, Vol. 5, No. s1, pp. s63–s66 (2021).
- [23] 佐藤翔輔, 今村文彦: 2018 年西日本豪雨災害における「#救助」ツイートの実態: 2017 年 7 月九州北部豪雨災害との比較分析, 自然災害科学, Vol. 37, No. 4, pp. 383–396 (2019).
- [24] 消防庁国民保護・防災部防災課: 市町村の災害対策本部機能の強化に向けて ～防災情報システム活用事例集～, <http://www.fdma.go.jp/neuter/topics/pdf/20170703.pdf> (2017). (2023/3/1 確認).
- [25] 末吉優, 関洋平: 音楽のジャンルと印象を用いた VOCALOID クリエータの検索, 人工知能学会論文誌, Vol. 32, No. 1, pp. WII-K.1–12 (2017).
- [26] 杉原健一郎, 石野航平, 松下光範: 災害情報を対象とした意思決定支援システムの検討: 複数人による協調的情報トリアージを目指して, 第 4 回 ARG WEB インテリジェンスとインタラクション研究會予稿集, pp. 75–80 (2014).
- [27] 宋晨潔, 藤代裕之: 災害時における信頼性の高い救助要請の見つけ方～西日本豪雨「救助」ツイートの検証～, 電子情報通信学会技術研究報告, Vol. 118, No. 439, pp. 7–12 (2019).
- [28] 宋晨潔, 藤代裕之: 救助要請ツイートの特徴の検証 —令和 2 年 7 月豪雨を対象に—, 電子情報通信学会技術研究報告, Vol. 120, No. 166, pp. 18–23 (2020).
- [29] 総務省情報通信政策研究所: 各種サービス（ソーシャルメディア系サービス/アプリ、ニュースサービス等）の利用率等, 令和 2 年度情報通信メディアの利用時間と情報行動に関する調査報告書, 総務省, pp. 66–69 (2021).
- [30] 高橋暁子: ついに「Twitter 離れ」が始まった…イーロン・マスクを悩ませる「日本人の大量離脱」の現実味, PRESIDENT Online, <https://president.jp/articles/-/63880> (2023). (2023/3/1 確認).
- [31] 鳥海不二夫: 無料 TwitterAPI のサービス停止に関わる阿鼻叫喚, Yahoo!ニュース, <https://news.yahoo.co.jp/byline/toriumifujio/20230203-00335524> (2023). (2023/3/1 確認).
- [32] 日本放送協会: 長野県 ツイッターの救助要請収集で約 50 件救助に, <https://www.nhk.or.jp/politics/articles/lastweek/25652.html> (2019). (2023/3/1 確認).
- [33] 土方嘉徳: ソーシャルメディア論—行動データが解き明かす人間社会と心理—, サイエンス社 (2020).
- [34] 藤代裕之, 松下光範, 小笠原盛浩: 大規模災害時におけるソーシャルメディアの活用—情報トリアージの適用可能性, 社会情報学, Vol. 6, No. 2, pp. 49–63 (2018).
- [35] 松浦立樹: フォロワー数 318 万人の「地震速報」bot「今後の運用は難しい—Twitter API 有料化の余波大きく, ITmedia News, <https://www.itmedia.co.jp/news/articles/2302/03/news178.html>

(2023). (2023/3/1 確認).

- [36] 森野穰, 松下光範: 災害情報収集におけるエンタテインメントコンテンツによるコンタミネーションの調査, 情報処理学会研究報告, Vol. 2022-EC-65, No. 33, pp. 1-2 (2022).
- [37] 安尾萌, 北村茂生, 松下光範: 災害発生時における被害状況把握を目的とした情報共有システムの基礎検討, 電子情報通信学会 HCG シンポジウム 2018 論文集, No. B-6-2 (2018).
- [38] 横田由佳: 看護の果たすべき役割, 杏林医学会雑誌, Vol. 46, No. 4, pp. 295-299 (2015).

大規模 Web ニュース記事群における 論争発見と論点整理

Discovery and Organization of Controversial Topics in Large-Scale Web News Articles

藤原拓己^{1*} 吉岡真治¹
Takumi Fujiwara¹ Masaharu Yoshioka¹

¹ 北海道大学大学院情報科学院

¹ Graduate School of Information Science and Technology, Hokkaido University

Abstract: In recent years, a huge number of Web news articles have been distributed by various media around the world. On the other hand, there is a background of increasing use of news aggregation services such as Google News, which can deliver articles personalized for individual interests, and to compensate for their shortcomings, It is desirable to provide means for users to comprehend topics from a bird's-eye view. The purpose of this study is to present information that promotes users' understanding of current affairs by discovering and organizing controversies that exist in articles on a specific topic.

1 はじめに

インターネットの発達によって全世界的に Web ニュース記事の配信が盛んであるが、同じ出来事を報じたものであってもニュースサイトごとに、話題を否定または賛成的どのような立場から発信するのかが異なる場合がある。一方でユーザーの立場においては、これらの話題に関するすべてのニュースサイトを閲覧することが困難であるため、Google ニュースや SmartNews などのニュースアグリゲーションサイトやサービスなどが提供する記事のみを閲覧する場合が増えている。しかしそのようなサービスは、近年フィルターバブルやエコーチェンバーという名称で指摘されるように、各ユーザーの興味に類する記事の提供しやすいといった性質があるため、それとは対照的に話題を俯瞰的に捉える手段を提供することができれば、世相を理解する有用な方法となる。

この問題に対し、我々は、世界中の様々なサイトから発信されるニュース記事群から、特定の話題に関する記事を抽出し、その賛否の分布からニュースサイトのスタンスを分析する研究が行なってきた [1, 2]。本研究ではこれらの研究を発展させ、これらのニュースサイトで扱うサブトピックを中心として、どのような「論争」が存在するのかを整理することで俯瞰的な理解につながる情報を提供することを目的とする。

論争という語は一般にある話題について対立する意見や主張による争いが生じている状況を意味するが、本研究においては特に、特定の話題をより細かく分割したサブトピックにおいて賛否が対立するような意見が量的に均衡して存在しているようなものについて「論争」が生じているものとして扱う。

本研究では、これまでの研究 [1, 2] と同様に、全世界で配信されるニュースのメタデータを配信しているサイトである The GDELT Project (以下、GDELT と略す)¹のデータを利用し、特定の話題に関する大規模 Web ニュース記事群を収集する。これらの特定の話題に関する記事群からサブトピックを構築し、賛否を示す意見の偏りなどを分析することによって、「論争」が起きているサブトピックの発見に加えて、その内部に存在する代表的な意見(論争文)を提示する方法を提案する。

また、GDELT の記事データからサブトピックの抽出と分析を行なった事前検討を行なった際に、反対または賛成の意見が大部分のサブトピックが構成されることが確認された。その中身について確認したところ、その記事群には同一性のある記事が多数存在し、その影響を軽減する必要がある。そのため本研究では冗長性の除去という観点で、記事群の前処理方法についても提案を行なう。

*連絡先：北海道大学大学院情報科学院
〒060-0814 北海道札幌市北区北 14 条西 9 丁目
E-mail: tfujiwara@kb.ist.hokudai.ac.jp

¹<https://www.gdeltproject.org/>

2 関連研究

SNS 上における論争の説明を試みる研究 [3] が存在する。これでは Twitter を対象として、論争の生じている話題についてその論争を説明するような最もよい要約ツイートを発見することを目的としている。その際、ツイート群を対立するようなスタンスを基準に二分する必要がある、そのために特定の話題を反映したようなハッシュタグをもつツイート群に対して、ツイート間のリツイート関係によって作成されるグラフから分割を求める手法 [4] が用いられており、本稿で扱う〈論争〉とはその性質において差異がある。

ニュース記事に限定しない Web 文書を対象とした関連研究として、ユーザーが入力した任意のクエリに対して、Web 上の文書情報を対象にその情報発信者や賛否意見などの情報を提供するために、WISDOM[5] というシステムや、クエリに関係するトピックについての文書群から、そこに含まれる言論を意味的に整理することで、上記の情報にくわえて多様な情報をユーザーに提供する言論マップ生成 [6] の手法が提案されている。これらは、Web 上の大規模な文書情報を対象とし、ある話題についての俯瞰的な理解につながる情報をユーザー提供する手法という意味において本研究と類似性があるが、広範に論争のあるようなサブトピックを探すことは行っていない。

3 ニュース記事群からの論争発見と論点整理

3.1 概要

本研究では Web ニュース記事の収集のために、世界中のニュース記事からその記事中に現れる固有名詞などのメタデータを付与して公開されている GDELT の GKG というデータベースを参照することによって行なう。また、本文の抽出には、Newspaper3k²を用いた。さらにそこに存在する記事本文をトピックごとのクラスターへ分割する際には Top2Vec[7]を用いる。Top2Vec では文書の分散表現を得る手法 [8] の実装である Doc2vec³などによって入力文書とそこに含まれる語の分散表現を作成し、UMAP[9]によって次元削除を行なったのち、DBSCAN[10]やDENCLUE[11]といったクラスタリング手法を性能改善したものである HDBSCAN[12]を適用し、分散表現の密集箇所を検出することでトピックの発見を行なう。各トピッククラスターの重心をトピックベクトルとして、これと任意の文書のコサイン類似度がその文書のトピックに対するスコアとして算出さ

れる。また Top2Vec ではトピッククラスター同士を統合することで任意の個数にそのクラスター数を削減する階層的トピックの機能を提供する。

そして、〈論争〉の発生状況を考慮するために、記事文書に賛否情報を付与する必要があるが、これには畳み込みニューラルネットワークを用いた手法 [13]を用いている Python の自然言語処理パッケージ Stanza の Sentiment Analysis 機能を用いる。

これらの要素を用いて、論争発見と論点整理に有用なクラスターを構築することで、〈論争〉の内容を具体化する。

3.2 記事群に対する前処理

先にも述べたように、GDELT から取得される記事群には、同一性のある記事が複数存在するという問題がある。本章では、そのような記事の取り扱いとサブトピック分析を行なう際の関連キーワード記事群の選択や、収集した記事から意見性の高い文章を選択するための手法について紹介する。

3.2.1 ニュースアグリゲーションサイトの削除

ニュースサイトのうち、様々なスタンスをもつサイトから記事を寄せ集めたものであるニュースアグリゲーションサイトは、記事に意見性が存在しないと考えられるため、収集された記事から取り除いた。また、“freerepublic.com”などといったフォーラムサイトも同様のものとみなした(表 1)。

GKG から収集された記事群からこれらのサイト発行のものを取り除いた記事群を分析対象記事群 A とした。

表 1: アグリゲーションサイトとみなしたサイト
サイト名

au.finance.yahoo.com	au.news.yahoo.com
au.tv.yahoo.com	ca.finance.yahoo.com
ca.news.yahoo.com	ca.sports.yahoo.com
ca.style.yahoo.com	finance.yahoo.com
freerepublic.com	in.finance.yahoo.com
in.news.yahoo.com	news.yahoo.com
nz.finance.yahoo.com	nz.news.yahoo.com
ph.news.yahoo.com	sg.finance.yahoo.com
sg.news.yahoo.com	uk.finance.yahoo.com
uk.news.yahoo.com	uk.style.yahoo.com
www.yahoo.com	www.msn.com

²<https://newspaper.readthedocs.io/>

³<https://radimrehurek.com/gensim/models/doc2vec.html>

3.2.2 記事群の冗長性削減

ニュースサイトには通信社や系列メディアが存在し、すなわちニュース記事には配信元・配信先で同一性のある記事が配信される場合がある。このような親子関係のモデルが生じているときに、本研究では配信元のサイトを Provider Site と呼ぶこととし、記事群から、そのようなサイトの抽出を試みた。

Provider Site が求められたならば、重複している記事群が存在する場合に、Provider Site の発行した記事の1つを意見を発した代表記事と定めることで、記事の重複性排除に代えて、サイトごとの意見を分析する際に、サイト別記事群のサイズを大きくすることによってより確実な分析が可能となるほか、同一のサイト内同一記事によって生じる重複性も取り除くことができる。分析対象記事群 A について、記事同士の同一性に注目したグラフ $G_A = (A, E_A)$ を作成した。記事同士が同一性をもつ条件として、タイトルまたは本文(の一部)が等しいということに注目した。本文については、ニュース記事上に掲載されている記事本文中には記事の内容を示す文字列に加えて、それぞれの記事に固有な広告や警告文が含まれる場合があるため、以下のように記事 $a_i \in A$ を部分文字列に分割した文字列集合 L_{a_i} を考慮した。ただし Printable は入力された文字列について英文印字可能な文字⁴を同じ順で抜き出した文字列を返す関数で、PSplit は、ピリオド(“.”)で区切った文字列集合を返すものである。

$$L_{a_i} = \{\text{Printable}(t) \mid t \in \text{PSplit}(a_{i,\text{text}})\}. \quad (1)$$

記事同士の比較対象としての部分文字列 $l_{a_i,j} \in L_{a_i}$ を十分大きいものに限定するため、および G_A 上の辺 E_A 作成時の計算量を削減する目的で部分文字列から一定の文字列長 t_{len} をもつものを L'_{a_i} として採用した。

$$L'_{a_i} = \{l_{a_i,j} \in L_{a_i} \mid |l_{a_i,j}| \geq t_{\text{len}}\}. \quad (2)$$

ここで、 t_{len} は $l_{a_i,j}$ の平均長の整数部分とした。

$$t_{\text{len}} = \left\lfloor \frac{\sum_i \sum_j |l_{a_i,j}|}{\sum_i |L_{a_i}|} \right\rfloor \quad (l_{a_i,j} \in L_{a_i}). \quad (3)$$

E_A は無向辺であり、

$$P_{\text{idn}}(a_i, a_j): a_{i,\text{title}} = a_{j,\text{title}} \vee |L'_{a_i} \cap L'_{a_j}| \geq t_{\text{elem}} \quad (4)$$

$$E_A = \{(a_i, a_j) \in A^2 \mid i < j \wedge P_{\text{idn}}(a_i, a_j)\} \quad (5)$$

つまり、相異なる記事同士を比較したときに、そのタイトル文字列が等しい、または L'_{a_i} と L'_{a_j} とで部分文字列が t_{elem} 個以上共通する⁵場合に辺が定義される。このように、グラフ上で同じ辺を共有している場合にそれらの記事同士に同一性があるものとする。ただし t_{elem} は以下のようにした。

$$t_{\text{elem}} = \max \left\{ \min \left\{ |L'_{a_i}|, |L'_{a_j}| \right\}, 1 \right\}. \quad (6)$$

Provider Site は複数のオリジナル記事を作成し、配信先のサイトはそれらから選択的に記事を受け取り、掲載していると考えられる。つまり、記事同一性グラフ G_A 上で考えた場合、Provider Site の方がより複数の連結成分の記事の発行元としてみられるものとして G_A から Provider Site を抽出する方法を考案した (Algorithm 1)。これを行なうためにまず、 G_A を連結成分ごとに分割した部分グラフ集合 $C_A = \{G_1, G_2, \dots, G_n\}$ について、 G_i に含まれるサイト群

$$S_i = \{s \mid a_j \in G_i, s = a_{j,\text{site}}\} \quad (7)$$

を考える必要がある。

さらに、 $G_i \in C_A$ にただ1つ対応する記事 $\alpha_i \in G_i$ を代表記事とする。代表記事群

$$A_{\text{repr}} = \{\alpha_1, \alpha_2, \dots, \alpha_n\} \quad (8)$$

を決定するために、得られた Provider Site 列を S_{pro} を用いて Algorithm 2 を実行した。これでは各連結成分 G_i が $|G_i| > 1$ の場合に、Provider Site 列に存在していて、より列の先頭の位置にあるサイトが発行元となるような記事を代表記事 α_i として採用する。なお Algorithm 1 より、 $|G_i| > 1$ のとき $a_{j,\text{site}} \in S_{\text{pro}} (a_j \in G_i)$ となるような記事 a_j が存在することが保証される。

Algorithm 1 Provider Site 列を抽出するアルゴリズム

```

1: procedure CREATE-PROVIDER-SITE-RANKING( $C_A$ )
2:   Initialize  $h_{\text{freq}} \triangleright$  A hash of a site and its frequency
3:   for all  $s \in S$  do
4:      $C_s \leftarrow \{G_j \in C_A \mid s \in S_j\}$ 
5:   end for
6:   for all  $G_i \in C_A$  do
7:     if  $|G_i| > 1$  then
8:        $s_{\text{max}} \leftarrow \text{SELECT-RANDOM-ONE} \left( \underset{t}{\operatorname{argmax}} |C_t| \right)$ 
9:        $h_{\text{freq}}(s_{\text{max}}) \leftarrow h_{\text{freq}}(s_{\text{max}}) + 1$ 
10:    end if
11:  end for
12:   $S_{\text{pro}} \leftarrow \text{GET-KEYS}(h_{\text{freq}})$ 
13:   $\text{SORT-DESC-BY-FREQ}(S_{\text{pro}}, h_{\text{freq}})$ 
14:  return  $S_{\text{pro}}$ 
15: end procedure

```

⁴ここでは、英アルファベット大文字・小文字、アラビア数字および記号(!"#\$%&'()*+,-./:;<=>?@[^_`{|}~), 空白とした。

⁵文字列比較の際は計算量削減のため、実装において C++ 標準ライブラリの `std::hash<std::string>` 構造体で作成したハッシュ値による比較を行なった。

Algorithm 2 代表サイト群 A_{repr} を決定するアルゴリズム

```

1: procedure EXTRACT-REPR-ARTICLES( $C_A, S_{\text{pro}}$ )
2:   Initialize  $A_{\text{repr}} = \emptyset$ 
3:   for all  $G_i \in C_A$  do
4:     if  $|G_i| = 1$  then
5:        $A_{\text{repr}} \leftarrow A_{\text{repr}} \cup \{a_j\} \ (a_j \in G_i)$ 
6:       continue
7:     end if
8:     for all  $s \in S_{\text{pro}}$  do
9:       for all  $a_j \in G_i$  do
10:        if  $a_{j,\text{site}} = s$  then
11:           $A_{\text{repr}} \leftarrow A_{\text{repr}} \cup \{a_j\}$ 
12:          go to 15
13:        end if
14:      end for
15:    end for
16:  end for
17: end procedure

```

3.3 論争発見と論点整理

Top2Vec は、代表的な話題を見つけることができる一方、話題に関する記事数のばらつきが大きく、話題が小さくなりすぎると、反対または賛成の意見がほとんどのサブトピックなどが構成されることがあった。本研究ではこれらのサブトピックを賛否の異なる類似したサブトピックを併合していくことで、共通の話題に対する賛否が存在するようなサブトピックの作成を行なうことで、「論争」の内容を具体化する方法を提案する。

3.3.1 入力対象記事文書の決定

Top2Vec への入力対象とする記事群 A_{inp} を以下のよう定めた。

$$A_{\text{inp}} = A_{\text{repr}} \cap A_{\text{kwd}}. \quad (9)$$

ただし、 A_{kwd} は任意のキーワードに関連する記事群であり、 A に含まれる記事であって、記事の参照元の各 GKG のデータにおいてそのフィールドのいずれかにキーワードが部分文字列として含まれる場合に、その記事は A_{kwd} に含まれる。

そして、文章の文単位への分割には高速動作を期待し、Python の自然言語処理ライブラリである spaCy の Tokenizer の機能を用いた。また、各文には Stanza の Sentiment Analysis 機能により、賛否情報を示す 0, 1, 2 のいずれかの値を付与した。

ここで、分割された文について意味のあるものを扱うため、文の構成トークン数が 10 個以上、かつ文に動詞が含まれるようなものを入力対象とした。各トークンへは Stanza によって分割するとともに POS タグを付与したが、このうち“VERB” (動詞) および“AUX” (助動詞) のタグをここでは動詞と定義して扱った。

3.3.2 Top2Vec によるトピックモデリング結果の利用

記事 $a_i \in A_{\text{inp}}$ に含まれる、3.3.1 で得られた文を含む文書集合 $\mathcal{D}_{a_i} = \{D_{a_i,1}, D_{a_i,2}, \dots, D_{a_i,n}\}$ について、すべての入力文集合 $\mathcal{D}_{\text{inp}} = \bigcup_i \mathcal{D}_{a_i}$ を Top2Vec に入力する⁶ことで、文あるいはそこに含まれる単語による分散表現を得て、さらにそれをトピックごとに分割するクラスタリングを行なった。ただし、 $D_{a_i,j}$ は文 $d_{a_i,j}$ とその感情値 $c_{a_i,j}$ 、サイト名 $s_{a_i,j}$ を含む列 $(d_{a_i,j}, c_{a_i,j}, s_{a_i,j})$ である。

Top2Vec で $\mathcal{D}_{\text{inp}}$ を対象にトピックモデリングを行ない、トピックを示すクラスターを含む、下位トピッククラスター群 $\mathcal{T}_{\text{lower}} = \{T_i, T_j, \dots\}$ ($T_i = \{D_{a_k,l}, D_{a_m,n}, \dots\}$) を作成した。さらに Top2Vec のトピック削減機能により、 $|\mathcal{T}_{\text{upper}}|$ を任意の値に定めたうえで $|\mathcal{T}_{\text{upper}}| < |\mathcal{T}_{\text{lower}}|$ となるような上位トピッククラスター群 $\mathcal{T}_{\text{upper}} = \{T_k, T_l, \dots\}$ を得た。ここで、 $\mathcal{T}_{\text{upper}}$ の T_i が $\mathcal{T}_{\text{lower}}$ の T_j を包含する場合、 T_j は T_i のサブトピックであるとみなし、サブトピック関係 R_{sub} を定義した。

$$R_{\text{sub}} = \{(T_i, T_j) \in \mathcal{T}_{\text{upper}} \times \mathcal{T}_{\text{lower}} \mid T_i \supseteq T_j\}. \quad (10)$$

3.3.3 上位トピッククラスターに関する代表的な賛否文の発見

論争文の発見を行なう前に、任意の上位トピッククラスターに関して賛否いずれかを代表するような文の発見を行なった。

T_i のサブトピッククラスター群 $\mathcal{T}_i = \{T_j \mid (T_i, T_j) \in R_{\text{sub}}\}$ について、ブトピッククラスターに存在している文書をサイト s ごとに分割した集合

$$T_{i,s} = \left\{ D_j \in \bigcup_{T \in \mathcal{T}_i} T \mid s \in D_j \right\} \quad (11)$$

を用いて、否定派サイト群 $S_{i,\text{neg}}$ および賛成よりサイト群 $S_{i,\text{pos}}$ を得た。

$$S_{i,\text{neg}} = \{s \mid n_{\text{neg}}(T_{i,s}) > n_{\text{pos}}(T_{i,s})\} \quad (12)$$

$$S_{i,\text{pos}} = \{s \mid n_{\text{pos}}(T_{i,s}) > n_{\text{neg}}(T_{i,s})\}. \quad (13)$$

これより、否定的文書群および賛成的文書群を求めた。否定的文書群 $\mathcal{D}_{i,\text{neg}}$ は、

$$\mathcal{D}_{i,\text{neg}} = \left\{ D_j \in \bigcup_{T \in \mathcal{T}_i} T \mid s, c \in D_j, s \in S_{i,\text{neg}} \wedge c = 0 \right\} \quad (14)$$

⁶Top2Vec (Ver. 1.0.26) の実装では、文書として文字列が対象のため、実装上、 $(c_{a_i,j}, s_{a_i,j})$ が対応づけられた $d_{a_i,j}$ を入力した。

とし、賛成的文書群 $\mathcal{D}_{i,\text{pos}}$ は (14) 式において条件を $c = 2$ としたものと定義した。

文書 D に対する文書ベクトル v_D を用いて、 $T_{i,\text{neg}}$ に属する文書ベクトルの重心を上位トピック T_i ($\in \mathcal{T}_{\text{upper}}$) に対する否定的文書群重心 $v_{i,\text{neg}}$ とし、

$$v_{i,\text{neg}} = \frac{\sum_{D \in \mathcal{D}_{i,\text{neg}}} v_D}{|\mathcal{D}_{i,\text{neg}}|} \quad (15)$$

賛成的文書群重心も同様に定義した。ただし文書ベクトルは次元削減前の次元空間のものである。

否定的文書群重心 $v_{i,\text{neg}}$ に対してそのベクトルが最もコサイン類似度の高くなるように文書群 $\mathcal{D}_{i,\text{neg}}$ を並べたものを上位トピック T_i ($\in \mathcal{T}_{\text{upper}}$) に対する否定的代表文群の一覧とした (ただし一覧のうち、文書に同じサイトまたは文が重複して現れる場合はその類似度が最も大きいものを残し、重複分を削減した)。賛成的代表文の一覧も同様に求めて。

3.3.4 賛否の均衡を考慮したトピッククラスターの再構成

T_i ($\in \mathcal{T}_{\text{upper}}$) のサブトピッククラスター群においてクラスター同士の併合を行なうことで、賛否文数の比率がなるべく均等となるようなクラスターの作成方法を考案した (Algorithm 3)。クラスター T の賛否特性として行 8 にあるように評価値 $p(T)$ を定義した。 $p(T)$ は賛成的な文書数 $n_{\text{neg}}(T)$ に比して否定的な文書数 $n_{\text{pos}}(T)$ がどの程度存在するののかという極性を文書数 $|T|$ によって正規化したものである。

$$p(T) = \frac{n_{\text{neg}}(T) - n_{\text{pos}}(T)}{|T|}. \quad (16)$$

クラスターの統合を繰り返し行なうと、極性の大きく偏った 1 つのクラスターへと肥大化する事象が生じることを確認したため、統合されたクラスターのサイズが t と $\mathcal{T}_{\text{sub}}$ によって定まる閾値以上となった場合にこれを取り除くようにした (行 14, 15)。

3.3.5 サイトごとの賛否状況を考慮した論争の検出

Web ニュース記事の配信元であるニュースサイトは、記事本文中の文書の内容の発言者であると考えられるため、サイトごとの否定または賛成派グループを手がかりとして論争の検出を行なう。

T_i ($\in \mathcal{T}_{\text{upper}}$) の内部でのように再構成されたサブトピッククラスター群 $\mathcal{U}_i = \{U_{i,1}, U_{i,2}, \dots, U_{i,n}\}$ について、 $U_{i,j}$ 内で文書の発行元であるサイトが s であるような文書集合

$$V_{i,j,s} = \{D_{a_k,l} \in U_{i,j} \mid s \in D_{a_k,l}\} \quad (17)$$

を作成し、

$$\mathcal{V}_{i,j,\text{neg}} = \{V_{i,j,s} \mid n_{\text{neg}}(V_{i,j,s}) > n_{\text{pos}}(V_{i,j,s})\} \quad (18)$$

$$\mathcal{V}_{i,j,\text{pos}} = \{V_{i,j,s} \mid n_{\text{pos}}(V_{i,j,s}) > n_{\text{neg}}(V_{i,j,s})\} \quad (19)$$

として、比 $|\mathcal{V}_{i,j,\text{neg}}|/|\mathcal{V}_{i,j,\text{pos}}|$ に、 $U_{i,j}$ における賛否それぞれの文数の比 $n_{\text{pos}}(U_{i,j})/n_{\text{neg}}(U_{i,j})$ を乗算することで補正した値である

$$r(U_{i,j}) = \frac{n_{\text{pos}}(U_{i,j}) \cdot |\mathcal{V}_{i,j,\text{neg}}|}{n_{\text{neg}}(U_{i,j}) \cdot |\mathcal{V}_{i,j,\text{pos}}|} \quad (20)$$

を定義した。さらに

$$r'(U_{i,j}) = \begin{cases} r(U_{i,j}) & (r(U_{i,j}) \leq 1) \\ 1/r(U_{i,j}) & (r(U_{i,j}) > 1) \end{cases} \quad (21)$$

としたときに以下が真、偽となる場合にそれぞれ $U_{i,j}$ において論争が生じている、またはこれが生じていないと定義した。ただし $U_{i,j}$ に対して $r'(U_{i,j})$ が定義されない場合は $P_{\text{cont}}(U_{i,j})$ を偽とする。

$$P_{\text{cont}}(U_{i,j}): r'(U_{i,j}) \geq r_{\text{cont}}. \quad (22)$$

3.3.6 論争の起こっているクラスターからの論争文の発見

$P_{\text{cont}}(U_{i,j})$ が真となる $U_{i,j}$ について、論争文の発見を行なった。 $U_{i,j}$ から、否定ラベルが付与された文書群をそれらスコアの降順に並べたものを否定的な論争文列とした。賛成的な論争文列も同様に定義した。ただし、分析対象記事群 A の冗長性削減されない分も考慮し、それぞれの論争文列において、文書はスコアの最も高いものを残すように重複分を削除した。

Algorithm 3 サブトピッククラスター再構成アルゴリズム

```

1: procedure REMODEL-SUBTOPIC-CLUSTERS-IN-
   TOPIC( $T_i, t$ )
2:   Initialize  $\mathcal{T}_{\text{popped}} = \emptyset$ 
3:    $\mathcal{T}_{\text{sub}} \leftarrow \{T_j \mid (T_i, T_j) \in R_{\text{sub}}\}$ 
4:    $\mathcal{T}_{\text{cur}} \leftarrow \mathcal{T}_{\text{sub}}$   $\triangleright$  Create a copy not a reference
5:   repeat
6:     for all  $T \in \mathcal{T}_{\text{cur}}$  do
7:        $(n_{\text{neg}}(T), n_{\text{pos}}(T)) \leftarrow$ 
         COUNT-SENTIMENT-LABELS( $T$ )
8:        $p(T) \leftarrow (n_{\text{neg}}(T) - n_{\text{pos}}(T))/|T|$ 
9:     end for
10:     $T_{\min} \leftarrow \text{SELECT-RANDOM-ONE} \left( \underset{T \in \mathcal{T}_{\text{cur}}}{\text{argmin}} p(T) \right)$ 
11:     $T_{\max} \leftarrow \text{SELECT-RANDOM-ONE} \left( \underset{T \in \mathcal{T}_{\text{cur}}}{\text{argmax}} p(T) \right)$ 
12:     $T_{\text{merged}} \leftarrow T_{\min} \cup T_{\max}$ 
13:     $\mathcal{T}_{\text{cur}} \leftarrow \mathcal{T}_{\text{cur}} \setminus \{T_{\min}, T_{\max}\}$ 
14:    if  $|T_{\text{merged}}| \geq \left( \sum_{U \in \mathcal{T}_{\text{sub}}} |U| \right) / t$  then
15:       $\mathcal{T}_{\text{popped}} \leftarrow \mathcal{T}_{\text{popped}} \cup \{T_{\text{merged}}\}$ 
16:    else
17:       $\mathcal{T}_{\text{cur}} \leftarrow \mathcal{T}_{\text{cur}} \cup \{T_{\text{merged}}\}$ 
18:    end if
19:  until  $|\mathcal{T}_{\text{cur}}| \leq 1$ 
20:  return  $\mathcal{T}_{\text{popped}} \cup \mathcal{T}_{\text{cur}}$ 
end procedure

```

4 実験

3.2 および 3.3 の手法を用いて、具体的な条件設定のもとで実験を行なった。

4.1 記事群に対する前処理

収集する記事については、“Biden”, “Putin”, “Ukraine” または “Zelensky” のいずれかをキーワードとして含む英文で記述されたもので、2022 年 1 月 1 日から 4 月 30 日の間に GKG のデータベースにおいて集計されたレコードから取得されるものを対象とした。それらよりニュースアグリゲーションサイトの発行した記事を取り除いた分析対象記事群 A の個数 $|A|$ は 477,689 であった。

また、 A から抽出された Provider Site 列は初項から、 $(\text{site}, h_{\text{freq}}(\text{site})) = (\text{www.europesun.com}, 2,100)$, $(\text{www.breitbart.com}, 1,025)$, $(\text{www.newsbusters.org}, 904)$, $(\text{apnews.com}, 875)$, $(\text{www.cnn.com}, 818)$, ... となった (参考として h_{freq} の値も示した)。

4.2 論争発見と論点整理

4.2.1 分析対象記事の決定

2022 年の出来事として 2 月に起きたロシアによるウクライナ侵攻がある。このことに際してウクライナ側の対応が注目され、当時のウクライナ第 6 代目大統領のウォロディミル・ゼレンスキーの発言や振る舞いがメディアに取り上げられた。実験においては、この出来事およびゼレンスキーについての記事群を対象とすることを目的として、キーワードとして “Zelensky” を含む記事群 A_{kwd} をもとに入力記事群 A_{inp} を決定した。各記事数 $|A_{\text{repr}}|$ $|A_{\text{kwd}}|$ $|A_{\text{inp}}|$ はそれぞれ 224,357, 13,913, 12,191 件で、入力記事文書群の個数 $|D_{\text{inp}}|$ は 568,718 件であった。

4.2.2 下位および上位トピッククラスターの取得

Top2Vec に D_{inp} を入力し、下位トピッククラスター群 $\mathcal{T}_{\text{lower}}$ を得た (表 2)。要素数 $|\mathcal{T}_{\text{lower}}|$ は 6,639 であった。さらに、 $|\mathcal{T}_{\text{upper}}|$ が 100 となるように Top2Vec の階層的トピック削減を行ない $\mathcal{T}_{\text{upper}}$ を得た (表 3)。Top2Vec は Ver. 1.0.26 の実装⁷を用いて、また実行の際、各種パラメータはデフォルトとした。

⁷<https://github.com/ddangelov/Top2Vec/releases/tag/1.0.26>

4.2.3 サブトピッククラスターの取得

$T_4, T_9 (\in \mathcal{T}_{\text{upper}})$ のサブトピックを実験対象としたため、それぞれについてのサブトピッククラスターを求めた (T_4 についての結果は表 4)。ただし、それぞれのトピックのサブトピッククラスター T_i について (20) 式と同様に $r(T_i)$ を計算し、その降順に並べた。この値が大きいほど、否定的な意見をもつニュースサイトがクラスターにより多く含まれるようなサブトピックとなっている。

T_4 については、price, inflation, consumers などがキーワードとしてあることから、市場や物価に関連するトピックで、 T_9 は、gop, republican, democrats などがみられるので、米国の政党に関するものであることが考えられる (表 2)。

4.2.4 サブトピッククラスターに関する代表的な賛否文の発見

以下ではまず、 $T_4 (\in \mathcal{T}_{\text{upper}})$ を対象として実験を行なう。 T_4 に関する否定的代表文一覧および賛成的代表文一覧を求めた。それによると主な否定的論点としては、「エネルギーや食糧価格の高騰」や「気候変動の抑制のためにエネルギー削減が必要」なことなどについてが挙げられ、賛成的論点としては、「小麦や金属類の価格高騰」や「米国での雇用増加」などがあった。

4.2.5 論争の起こっているクラスターからの論争文の発見

Algorithm 3 によって再構成したサブトピッククラスター群 $\mathcal{U}_4 = \{U_1, U_2, U_3, U_4\}$ を得た (表 5)。このうち論争が生じているとされた U_2 について、否定的または賛成的な論争文の上位 10 件を調べた。

得られた否定的な論争文としてはほとんどが「インフレのともなうエネルギーや物価の高騰や供給不安」についてであった。 T_4 の否定的代表文においても、主にエネルギーや食糧の価格高騰についてが論点となっていたため、発見された否定的論争文においては上位のトピックとの関連性がみられた。

賛成的論争文についても、同様のコストの高騰について述べているものがみられる (この情報だけでは賛否どちらの意見として発せられたかが判断できないが、感情分析の際に「物事が上昇する、高止まりする」といった表現が賛成的であると判定された可能性がある)。一方で、上位トピックの賛成的代表文には存在していなかった、ロシアが原油の輸出を再開したことで原油供給の懸念が緩和された旨の論争文も存在しており、これについては否定的論争文にもあったエネルギーの供給不安といった内容の対となる。

以上により、同じイベントに対して、賛否両面からの論争文が得られることがわかった。しかし、上記の話題はあくまで事実を報じた文であって、論争の対象としては第三者の感情をとまなうような意見が期待されるため、提示される論争文とはその性質において不適当であると考えられる。

4.2.6 他トピックを扱った追加実験

$T_9 \in \mathcal{T}_{\text{upper}}$ に関して、否定的代表文には、「Cawthorn 議員が Zelenskyy に対して過激な発言をした」ことや「そのことが他の議員から批判されている」ことが得られ、および賛成的代表文として「下院の選挙が盛況である」ことや「Cawthorn がノースカロライナ州で最も有名な人物である」ことなどが得られた。

そして T_9 のサブトピックを再構成したトピッククラスター群 $\mathcal{U}_9 = \{U_1, U_2, U_3, U_4, U_5\}$ を作成し、このうち論争が生じていると判定された U_4 について否定的または賛成的な論争文の発見を行なった。否定的な論争文としては、否定的代表文と同様のものが多く見られたが、賛成的な論争文について、「Cawthorn が 25 歳で共和党下院予備選に勝利した」ことや「Cawthorn が 1 期目にして多くの注目を集めた」などといった内容がみられた。

この結果より、否定または賛成的代表文でも Cawthorn について異なる賛否述べていた内容について扱われていたが、得られた否定的または賛成的代表文では、同氏についてより詳細な形で賛否両面からの論点が現れたと考えられる。

4.2.7 考察

上位トピック T_4 はキーワードとして price, inflation, consumers など、事柄や概念についてで、 T_9 は gop, republican, democrats など、具体的な組織を表していると考えられる。 T_4 での結果としては、論争文として得られたものが物価の高騰などの出来事について述べたもので、論争として望まれるような性質を持ったものが発見されなかった。これに対して、 T_9 では、具体的な人物に対しての批判的あるいは賞賛的な論争文が発見され、これは論争として適当であった。具体的な組織や人物に関するトピックであれば、論争文として妥当なものが得られると考えられる。これは、特定の人物に関して述べた文は、その人物に向けた感情が表出しやすいためと考えられる。

以上より、トピックによっては論争が起こっていると判定されるものであっても、論争文の妥当性がないもの存在することが考えられる。

表 2: 下位クラスター $T_i \in \mathcal{T}_{\text{lower}}$ のトピック例

i	T_i のキーワード	$n_{\text{neg}}(T_i)$	$n_{\text{pos}}(T_i)$	$ T_i $
1	know, you, going, ...	130	14	927
2	nord, stream, pipeline, ...	281	17	865
3	stinger, javelin, armor, ...	236	16	853
4	olaf, scholz, chancellor, ...	162	24	795
5	iaea, mariano, grossi, ...	311	17	759
⋮				
6638	commitments, securing, independence, ...	0	0	10
6639	venezuelan, withdrawals, barrels, ...	0	8	8

表 3: 上位クラスター $T_i \in \mathcal{T}_{\text{upper}}$ のトピック例

i	T_i のキーワード	$n_{\text{neg}}(T_i)$	$n_{\text{pos}}(T_i)$	$ T_i $
1	know, think, going, ...	2,690	410	14,021
2	know, you, me, ...	2,702	655	12,375
3	kharkiv, city, miles, ...	2,418	133	9,195
4	prices, inflation, consumers, ...	2,702	731	8,808
⋮				
9	gop, republican, republicans, ...	2,394	413	8,369
⋮				
99	crystal, inch, waived, ...	879	281	3,866
100	kabul, debacle, taliban, ...	951	227	3,856

表 4: サブトピッククラスター $T_i \in \mathcal{T}_{\text{sub}}$ のトピック ($r(T_i)$ の降順)

i	T_i のキーワード	$n_{\text{neg}}(T_i)$	$n_{\text{pos}}(T_i)$	$ T_i $	$r(T_i)$
4645	minimum, wage, paycheck, ...	14	2	48	1.7143
4831	workforce, afford, care, ...	22	4	45	1.4545
1985	spike, prices, inflation, ...	31	7	98	1.4113
1039	chains, disruptions, chain, ...	42	6	128	1.3810
1770	communities, resources, rural, ...	12	3	103	1.3750
⋮					

表 5: $T_i \in \mathcal{T}_{\text{sub}}$ を再構成したサブトピッククラスター U_j ($t = 5, r_{\text{cont}} = 0.8$)

j	$r'(U_j)$	$n_{\text{neg}}(U_j)$	$n_{\text{pos}}(U_j)$	$ U_j $	$P_{\text{cont}}(U_j)$	U_j の構成要素 ($T_i \in \mathcal{T}_{\text{lower}}$)
1	0.5486	362	79	1,144	False	$T_{23}, T_{505}, T_{974}, T_{3032}, T_{3396}, T_{4680}, T_{5348}, T_{5672}, T_{6354}, T_{6517}$
2	0.8340	362	67	1,119	True	$T_{1039}, T_{1240}, T_{1770}, T_{1848}, T_{1915}, T_{1995}, T_{2666}, T_{2800}, T_{3248}, T_{4645}, T_{4754}, T_{4831}, T_{5757}, T_{6182}, T_{6290}, T_{6415}$
3	0.8034	335	70	1,004	True	$T_{319}, T_{506}, T_{781}, T_{2211}, T_{2583}, T_{3056}, T_{3530}, T_{3573}, T_{4548}, T_{5062}, T_{6387}$
4	0.5416	399	75	1,230	False	$T_{602}, T_{957}, T_{1333}, T_{1460}, T_{1748}, T_{1985}, T_{2242}, T_{2998}, T_{3142}, T_{3560}, T_{4353}, T_{4368}, T_{5847}, T_{6245}, T_{6265}, T_{6377}$

5 むすび

本研究では GDELT を参照し収集された Web ニュース記事群に対して、そこに存在する重複性を削減したうえで、Top2Vec によって作成されるトピッククラスター群を加工したものを利用して〈論争〉の発見および論点の整理を試みた。実験においては、個別具体的

なトピックについて論争文の発見と整理を行なったところ、本研究が望む〈論争〉の発見には扱うトピックの性質が影響することが示唆された。

今後の展望としては、Top2Vec で取得されるトピッククラスターをより記事文書の内容に即すようにそのサイズや分割を厳密に決定することや、文書への賛否情報付与の際に文の意見性と賛否情報の整合性をよりとれるような方法を考案することなどが考えられる。

参考文献

- [1] Masaharu Yoshioka, Myungha Jang, James Allan, and Noriko Kando. Visualizing polarity-based stances of news websites. *NewsIR@ ECIR*, 2079:6–8, 2018.
- [2] 立浪紀彦. ユーザの興味を反映したニュースサイトの多観点スタンス分析, 2020. 北海道大学大学院情報科学院修士論文.
- [3] Myungha Jang and James Allan. Explaining controversy on social media via stance summarization. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, page 1221–1224, New York, NY, USA, 2018. Association for Computing Machinery.
- [4] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.
- [5] 赤峯享, 宮森恒, 加藤義清, 中川哲治, 乾健太郎, 黒橋禎夫, and 木俣豊. Web 情報の信頼性検証のための情報分析システム wisdom. *言語処理学会第 14 回年次大会論文集 [本論文集]*, 2008.
- [6] 水野淳太, 渡邊陽太郎, 村上浩司, 松吉俊, 大木環美, 乾健太郎, and 松本裕治. 言論マップ生成技術の現状と課題. *言語処理学会第 17 回年次大会講演論文集*, pages 49–52, 2011.
- [7] Dima Angelov. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*, 2020.
- [8] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR, 2014.
- [9] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018.
- [10] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer, 2013.
- [11] Alexander Hinneburg and Hans-Henning Gabriel. Denclue 2.0: Fast clustering based on kernel density estimation. In *International symposium on intelligent data analysis*, pages 70–80. Springer, 2007.
- [12] Leland McInnes, John Healy, and Steve Astels. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205, 2017.
- [13] Ye Zhang and Byron Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015.

ネットショッピングにおける 他者に役に立つショップレビュー作成支援システム

Useful Shop Reviews Writing Support System for Online Shopping Users

西原陽子^{1*} 浅堀なぎさ¹ Junjie Shan²
Yoko Nishihara¹, Nagisa Asahori¹, and Junjie Shan²

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 立命館グローバル・イノベーション研究機構

² Ritsumeikan Global Innovation Research Organization

Abstract: This paper proposes a support system for writing useful shop reviews for online shopping users. We used Rakuten Ichiba's dataset, which contains shop reviews with evaluated scores. The proposed system has two functions: the review judgment function and the review example function. The review judgment function is implemented by learning reviews with fastText. The review example function could randomly display a review from highly scored reviews. The judgment function is used to support checking whether a review is useful, while the review example function supports finding what kind of information is necessary for an useful review. We experimented with the proposed system with 15 participants and found how the two functions affect writing useful reviews.

1 はじめに

多くのレビューが日々執筆され、他のユーザの意思決定に役立てられている。レストランのレビューや映画のレビュー、商品のレビューなど、様々なものに対してレビューが書かれている。レビュー内容の分析やレビューの極性分析など、執筆されたレビューに関する研究は多数存在する。例えば、レビューテキストの各文が何について書かれているのかを分類する研究 [7] や、レビューがどのメディアについての言及であるかを分類する研究 [4]、レビューの極性判定 [1]、トピック分析を行う研究 [3] などがある。

一方で、レビューの執筆を支援する研究にも注目が集まっている。レビュー執筆支援は文章執筆支援の研究の一種と考えられる。論文やレポートなどの技術文書の執筆を支援する研究や [6]、小説や随筆などの執筆を支援する研究 [5] などがある。レビューや口コミの執筆支援に関する研究としては、評価してほしい対象物や観点、それに使われる評価表現の与え方を変化させたときに、ユーザの書くレビューの内容がどのように変

化するかを明らかにしたものなどがある [2]。レビューの執筆が支援されると、より多くの良質のレビューを集めることができる。これはレビューを読むユーザにとっても、レビューを書いてもらった店舗や出品者側にとっても有益になる。

そこで本論文では、ネットショッピングを利用するユーザに向けて、他者に役立つショップレビューを書くための支援システムを提案する。他者に役立つレビューのデータとして、本研究では楽天市場のショップレビューのテキストと、レビューに対し他のユーザから「参考になった」と評価された数のデータを用いる。ショップレビューに対して他者から参考になったと評価されたということは、他者にとってその情報が役に立ったとみなせる。本研究では、機械学習のアプローチを用いて「参考になった」と評価されるレビューの特徴を学習し、その特徴に沿ったレビューの執筆を支援するシステムを提案する。

2 提案システム

はじめに提案システムの概要を説明する。提案システムは2つの機能を有する。1つはレビューが他者に

*連絡先：立命館大学情報理工学部
滋賀県草津市野路東 1-1-1
nishihara@fc.ritsumeikan.ac.jp

として役に立つものであるかを評価する「レビュー判定機能」であり、もう1つは役に立つと評価された中から内容が肯定的なものと否定的なものをランダムに表示する「例文ガチャ機能」である。ユーザは、自分が執筆したレビューを提案システムに入力し、レビューの判定機能を用いて執筆したレビューが役に立つかどうかを知る。レビューの内容を修正したいときには、例文ガチャ機能を用いて参考になるレビューを提示し、自らのレビューを修正する。2つの機能を組み合わせ、繰り返し使うことにより、他者にとって役立つショップレビューを完成することができる。

以下では、本研究で用いた楽天市場のデータセットについて述べ、その後、2つの機能の実装の詳細について説明する。

2.1 楽天市場のショップレビューのデータセット

本研究で用いたデータは楽天市場により提供されている市場店舗レビューデータセットである。用いたデータの投稿期間は、2018年12月から2019年12月分の13ヶ月分である。このデータの中には投稿者ID、店舗名、店舗ID、評価ポイント、レビュー本文、参考になった数、レビュー投稿日時が記されている。このうち「参考になった数」は、レビューを読んだユーザが「参考になった」ボタンを押した数である。本研究では「参考になった数」を用いて、役に立つレビューとそれ以外を分け、機械学習のための教師データを作成する。

教師データを作成するために、はじめにデータのクレンジングを行う。対象期間においてほとんどの投稿者は数回程度ショップレビューを書くだけであったが、一部の投稿者はより多くのショップレビューを書いていた。本研究では、同一の投稿者IDが対象期間内に上位3%の件数以上のレビューを投稿していた場合には、一般的な投稿者ではないとしてデータからレビューを削除した。

続いて「参考になった数」を用いて、役に立つレビューとそれ以外を分け、データセットを作成する。「参考になった数」が閾値以上のデータを役に立つレビューのデータ、すなわち正例のデータとし、反対に「参考になった数」が0のものを負例のデータとした。本研究では「参考になった数」が2以上のデータを役に立つレビューのデータとした。

2.2 レビューの判定機能

作成したデータセットを用いてレビューの判定機能を実装する。本研究では文書分類器の fastText¹を用い

¹<https://github.com/facebookresearch/fastText>

表 1: fastText のレビューの分類精度

ラベル\分類結果	正例	負例
正例	3943	657
負例	847	4481

て、レビューを学習することにより分類器を作成し、レビューの判定機能を実装する。データを学習用とテスト用に分け、学習用として正例データを 57,090 件と負例データを 39,628 件 fastText に与えて学習させた。テスト用として正例データと負例データを 4,600 件、5,328 件それぞれ用意し、分類精度の評価を行った。テストデータの分類結果を表 1 に示す。正例を正例と分類できる精度は 85.72%、負例を負例と分類できる精度は 84.1%であった。

2.3 例文ガチャ機能

例文ガチャ機能の実装方法について説明する。レビューの判定機能により執筆したレビューが他者の役に立つか立たないかはわかるが、どのように書けば役に立つと判定されるのかは以前としてわかりにくい。そこで、役に立つと評価されたレビューを示すことにより、レビュー修正の支援を行う。役に立つと評価されたレビューは、参考になった数が2以上のものとする。

レビューを書くときはショップを称賛する場合もあれば、苦情を申し立てる場合もあると考えられる。書きたいレビューの内容に応じて、例文が表示されることが望ましい。そこで、レビューを肯定的な内容と否定的な内容に分類²し、ユーザの選択に応じて表示する例文を変更する。2018年12月に投稿された中から、役に立つ肯定的なレビューを 162 件、役に立つ否定的なレビューを 162 件用意した。ユーザには肯定的なレビューか否定的なレビューかを選ばせ、用意した中からランダムにレビューを表示する。

2.4 機能が実装されたユーザインタフェース

2つの機能を備えたユーザインタフェースを実装した。ユーザインタフェースは、Python と Tkinter により実装された。図 1 に、提案システムのユーザインタフェースを示す。ユーザは、一番上のテキストボックスに執筆したレビューを入力する。下の「実行」ボタンを押すと、レビューの判定が行われ、判定結果が図 2 のように示される。「肯定的な例レビュー」か「否定

²今回は人手により分類を行ったが、レビューのポジティブネガティブを分類する機械学習による手法を今後検討する。

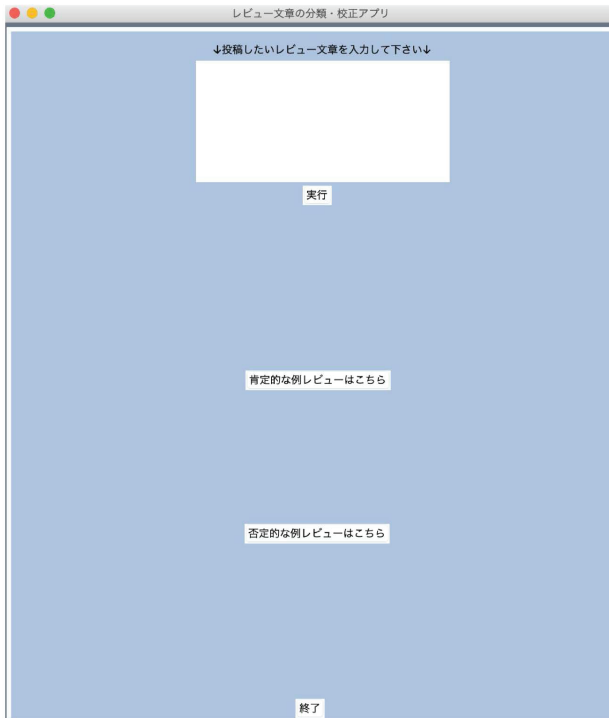


図 1: 提案システムのユーザインタフェース（上）、レビュー判定結果の例（中央）、および例文ガチャ機能の実行例（下）

「肯定的な例レビュー」ボタンを押すと、例文ガチャ機能が実行され、レビューの例文が示される。

3 評価実験

提案システムを評価する被験者実験を行った。被験者実験では、提案システムの2つの機能の有用性を評価した。

3.1 実験手順

はじめに実験手順を説明する。

1. 実験者は、被験者を募集する。
2. 被験者は、レビューを執筆するショップを選択し、レビューを執筆する。
3. 被験者は、レビューを修正する。与えられた場合はシステムを使用して修正し、与えられなかった場合は目視にて内容を確認し修正する。
4. 被験者は、2と3を3回行い、3つのレビューを作成する。

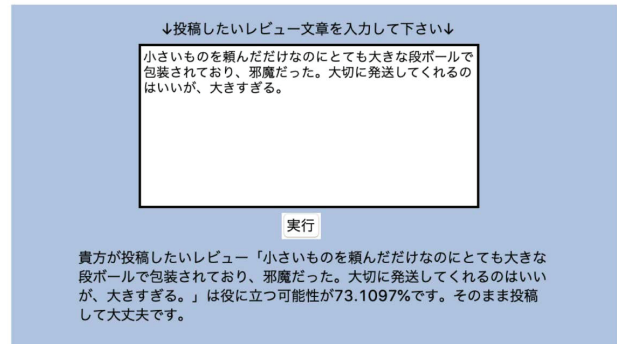


図 2: レビュー判定結果の例

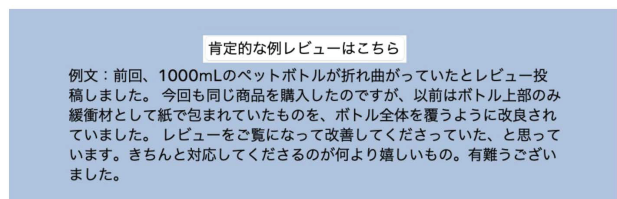


図 3: 例文ガチャ機能の実行例

5. 被験者は、完成した3つのレビューを実験者に提出し、アンケートに回答後、実験を終了する。

続いて、実験の詳細を説明する。被験者は、ネットショッピングの経験がある大学生、大学院生15名であった。年齢は20から24歳、男性が9名、女性が6名であった。

被験者は、レビューを修正する際、以下の3つの条件がそれぞれ課せられた。条件を課す順序は、被験者ごとに異なるように設定した。

- 何も使用しない：執筆したレビューの内容を目視で確認する。
- 判定機能のみ使用：レビューの判定機能だけが備わったシステムを用いて修正する。
- 提案システムを使用：レビューの判定機能と例文ガチャ機能の2つが備わったシステムを用いて修正する。

1つ目の「何も使用しない」の条件では、被験者はレビューを執筆するショップを決めた後、レビュー執筆のヒントとなる情報は一切閲覧せず、自力でレビューを執筆し、修正を行い、完成させる。

2つ目の「判定機能のみ使用」の条件では、被験者は1つ目と同様にレビューを執筆し、その後、レビュー判定機能を用いて自らが執筆したレビューが役に立つと判定されるかを確認する。役に立つと判定されれば修

正を終了し、役に立たないと判定されれば修正を繰り返す。ただし、判定機能を5回使用しても役に立たないと判定された場合は、それ以上は修正しない。

3つ目の「提案システムを使用」の条件では、被験者は1つ目と同様にレビューを執筆し、その後、レビュー判定機能を用いてレビューの判定結果を得る。続いて、書きたいレビューの内容に応じて例文ガチャ機能を用いて、例文を出力する。出力された例文を参考にし、レビューを修正する。例文ガチャ機能の使用回数に上限はないが、レビュー判定機能の使用回数の上限は2つ目と同様に5回とする。

被験者は、2つのアンケート項目に回答した。

- レビュー判定機能、例文ガチャ機能のどちらかがレビュー執筆に役立ったか（レビュー判定・例文ガチャ・どちらも役に立った・どちらに役に立たなかったの4択）
- 上記の理由（自由記述）

実験者は、レビュー執筆に要した時間を計測した。

- タイム1：ショップを選択してから、レビューを執筆し始めるまでの時間。
- タイム2：レビューを執筆し始めてから、レビューの作成が完了するまでの時間。

3.2 評価方法

評価は、2つの点から行った。なお、各条件で45個のレビューが得られ、3つの条件で135個のレビューが得られた。タイム1と2についても同様に、各条件で45個のタイムが得られ、3つの条件で合計135個のタイムが得られた。

- 各条件で得られた役に立つレビューの割合
- 各条件でのレビュー執筆までの平均時間（タイム1）とレビュー完成までの平均時間（タイム2）

3.3 実験結果

表2に、各条件で得られた役に立つレビューの数とその割合を示す。成功率は、提案システムを使用したときに65%、レビュー判定機能のみを使用したときに46%、何も使用しないときに28%であった。

続いて表3に、各条件でのレビュー執筆までの時間（タイム1）とレビュー完成までの時間（タイム2）の平均時間を示す。レビュー執筆までの時間は、提案システムを使用したときに17.06秒、レビュー判定機能のみを使用したときに15.18秒、何も使用しないとき

表 2: 各条件での役に立つレビュー数とその割合。提案システムを使用した場合で、ガチャ機能を使用せずにレビュー執筆を終了した分は含まれない。

条件	成功数	成功率
何も使用しない	13/45	28%
判定機能のみ使用	21/45	46%
提案システムを使用	23/35	65%

表 3: 各条件でのレビュー執筆までにかかる時間とレビュー完成までにかかる時間の平均

条件	執筆まで	完成まで
何も使用しない	13.64 秒	152.9 秒
判定機能のみ使用	15.18 秒	537.8 秒
提案システムを使用	17.06 秒	456.8 秒

が13.64秒であった。レビュー完成までの時間は、提案システムを使用したときに456.8秒、レビュー判定機能のみを使用したときに537.8秒、何も使用しないときに152.9秒であった。

最後に表4に、被験者のアンケート回答結果を示す。例文ガチャ機能を選択した人が53.3%で、どちらも役に立ったを選択した人が46.7%であった。

4 考察

本章では、レビュー判定機能と例文ガチャ機能について考察を行う。

4.1 レビュー判定機能に対する考察

表2では、「何も使用しない」ときは成功率が28%で「判定機能のみを使用」するときの成功率が46%であった。「判定機能のみを使用」するときの成功率が18%増加した。表5にレビューの判定機能が実際に使用され

表 4: レビュー執筆に役立った機能についての被験者アンケートの結果

選択肢	割合
レビュー判定機能	0%
例文ガチャ機能	53.3%
どちらも役に立った	46.7%
どちらも役に立たなかった	0%

表 5: 各条件でのレビュー判定機能の使用回数の平均

条件	レビュー判定機能の使用回数
判定機能のみ使用	3.867
提案システムを使用	3.044

表 6: 被験者が執筆した判定機能使用前後のレビューの例. 太字が修正された部分.

	レビュー
使用前	ズボンは質感も良く、色味はサイトの写真と変わりませんでした。匂いもついていませんでした。
使用后	ズボンは質感も良く、色味はサイトの写真と変わりませんでした。匂いもついていませんでした。 注文の翌日には発送された点がよかったです。

た回数の平均を示す。レビューの判定機能の使用回数の最大値は5回と設定していたが、被験者は「判定機能のみを使用」するときに平均3.8回使用していた。また、表3のレビュー執筆完了までにかかる時間は、「何も使用しない」ときは152.9秒で「判定機能のみを使用」するときは537.8秒であった。レビュー執筆完了までの時間が384.9秒増加したこと、レビューの判定機能を3.8回使用したことをあわせて考えると、被験者はレビューの判定機能を用いて自分のレビューの判定結果を得て、レビューに対して修正を行うことを繰り返すことができた。表6に、被験者が書いた判定機能使用前後のレビューの例を示す。判定機能を使用した後に、ショップの良いところをレビューに追加できている。以上のことから、レビューの判定機能を用いることにより、レビューの修正が促され、役に立つレビューを書く支援が行われることが明らかになった。

4.2 例文ガチャ機能に対する考察

表2では、「判定機能のみを使用」するときは成功率が46%で「提案システムを使用」するときは成功率が65%であった。「提案システムを使用」するとき、つまり例文ガチャ機能を使用することで成功率が19%増加した。表7に、被験者が書いた例文ガチャ機能の使用前後のレビューの例と、出力された例文の例を示す。例文ガチャ機能を使用した後に、ショップの懸念点をレビューに追加できている。例文ガチャ機能の平均使用回数は16.9回であったが、使用回数の最小値は0、最大

表 7: 被験者が執筆した例文ガチャ機能使用前後のレビューの例と出力された例文の一部. 太字が修正された部分.

	レビューまたは例文
使用前	使用前発送するのはとても早く、新しく買ったスマホを守れて良かったが、スマホケース自体の生地が安っぽい。
出力例文	(省略) 信頼していたショップなのでショックです！
使用后	発送するのはとても早く、新しく買ったスマホを守れて良かったが、スマホケース自体の生地が安っぽい。羊のショーンが描かれたスマホケースだったが、 しっかりとした本家のスマホケースを買うべきだと思う。

値は93で人によって使用回数が異なっていた。また、表3のレビュー執筆完了までにかかる時間は、「判定機能のみを使用」するときは537.8秒で「提案システムを使用」するときは456.8秒であった。レビュー執筆完了までの時間は81.0秒減少した。表5でもレビュー判定の機能の使用回数が3.867回から3.044回に減少した。使う機能は1つから2つに増えたが、レビュー判定機能と例文ガチャ機能をうまく併用することにより、レビュー執筆にかかる時間と判定回数を減少できている。以上のことから、例文ガチャ機能を用いることにより、より短時間でレビューの修正が行え、役に立つレビューを書く支援ができることが明らかになった。

5 おわりに

本論文では、ネットショッピングを利用するユーザーに向けて、他者に役立つショップレビューを書くための支援システムを提案した。提案システムは2つの機能を有し、1つは執筆されたレビューが役に立つかを判定する機能、もう1つは執筆に役立つ例文を表示する機能である。レビューの判定機能は、役に立つと評価されたレビューと評価されなかったレビューをfastTextを用いて学習し、レビューの分類器を作成することで実装した。例文を表示する機能は、役に立つと評価された中から肯定的な内容のレビューと否定的な内容のレビューを収集し、ランダムに表示することで実装した。

評価実験では被験者に提案システムの2つの機能を使ってショップレビューを書いてもらい、得られたレビューの判定結果と執筆にかかる時間から提案システムの評価を行った。実験の結果、レビューの判定機能と例文を表示する機能のどちらもがレビュー執筆を支

援することが明らかになった。今後は、提示する例文をユーザが書いているレビューの内容に応じて変化させることと、多様なレビューの執筆を支援できるシステムの構築に取り組んでいく。

謝辞

本研究の一部は、科研費(22H03708)と立命館グローバル・イノベーション研究機構の支援を受けました。記して謝意を申し上げます。本研究の一部に国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」を使用しました。(https://rit.rakuten.com/data_release/) 記して謝意を申し上げます。

参考文献

- [1] Hua Bai and Guojing Wang. Text sentiment analysis of hotel online reviews. In *Proceedings of the 5th International Conference on Computer Science and Software Engineering*, CSSE '22, p. 733 – 737, 2022.
- [2] Ryoya Suzuki, Ryosuke Yamanishi, and Yoko Nishihara. Framework for analysis and feedback of evaluation expressions in restaurant reviews. *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pp. 1–5, 2019.
- [3] Ye Yiran and Sangeet Srivastava. Aspect-based sentiment analysis on mobile phone reviews with lda. In *Proceedings of the 2019 4th International Conference on Machine Learning Technologies*, p. 101 – 105, 2019.
- [4] 藤居優奈, 韓毅弘, Junjie Shan, 西原陽子, 山西良典. 作品の展開メディアに対応したレビューの分類手法. 第8回コミック工学研究会, pp. 1–7, 2022.
- [5] 鈴木勘太, 杉本徹. Encoder-decoder モデルを用いた文章表現を豊かにする執筆支援システム. 日本感性工学会論文誌, 2022.
- [6] 伊藤拓海, 栗林樹生, 日高雅俊, 鈴木潤, 乾健太郎. Langsmith: 人とシステムの協働による論文執筆. 言語処理学会 第27回年次大会 発表論文集, pp. 1834–1839, 2021.
- [7] 山西良典, 藤岡寛子, 西原陽子. 擬似コーパスを用いた飲食店レビューの観点の自動分類. 人工知能学会論文誌, Vol. 36, No. 1, pp. 1–8, 2021.