

SNS における感情分析を用いた 関心の変化と要因推定

田中裕太^{†1} 佐野睦夫^{†1} 大井翔^{†2}

概要：近年，交友関係の構築や情報伝達などの様々な手段の一つとして SNS（ソーシャル・ネットワーキング・サービス）の利用が行われている。Twitter では，ユーザが日常生活で感じたことについてユーザそれぞれの見解を投稿している。そのため，口コミやレビューといった大衆の意見を得ることで製品やサービスの改善に役立てるための手段として利用されている。また，製品やサービスだけではなく，企業や行政が推進する何らかの取り組みに試みる際の関心の把握にも利用されている。本研究では，感情分析を用いることで Twitter のテキストから製品やサービス，取り組みに対する関心の変化の推定を行う。また，その関心の変化の手がかりとなる表現を抽出することでその要因の推定を行う。

キーワード：感情分析，形態素解析，トレンド

Estimating changes of interest and their factors using sentiment analysis on SNS

YUTA TANAKA^{†1} MUTSUO SANNO^{†1}
SHOJI OOI^{†2}

Abstract: In recent years, social networking services (SNS) have been used as one of various means of building friendships and communicating information. On Twitter, users post their own opinions about what they feel in their daily lives. For this reason, it is used as a means of improving products and services by gaining public feedback, such as word of mouth and reviews. It is also used not only for products and services, but also to understand interact when trying some kind of initiative promoted by companies and governments. In this study, we estimate changes in interact for products, services, and initiatives from Twitter text, and we estimate the factors by extracting expressions that provide clues to changes in interact.

Keywords: Sentiment Analysis, Morphological Analysis, Trend

1. はじめに

近年，交友関係の構築や情報伝達などの様々な手段の一つとして SNS（ソーシャル・ネットワーキング・サービス）の利用が行われている。多くの人が利用できる手段であると示すデータとして，図 1 の SNS 利用者数の推移のデータがある[1]。2012年度から 2016 年度までの 5 年間に於いて，日本における SNS の利用者率が 41.4% から 71.2% に増えている事がわかり，SNS は交友関係の構築や情報伝達の手段として普及していることがわかる。日本において普及している代表的な SNS には，Twitter（ツイッター）社が運営する Twitter[2]，Line（ライン）社が運営する Line[3]，Facebook（フェイスブック）社が運営する Facebook[4]，同じく Facebook 社が運営する Instagram（インスタグラム）[5]などがある。これらの SNS は，多くの利用者が交友関係の構築や情報伝達の手段として利用している。SNS には，交流のある個人や複数人の間における話し合いの場を設ける機能だけでなく，不特定多数のユーザに向けてユーザ自

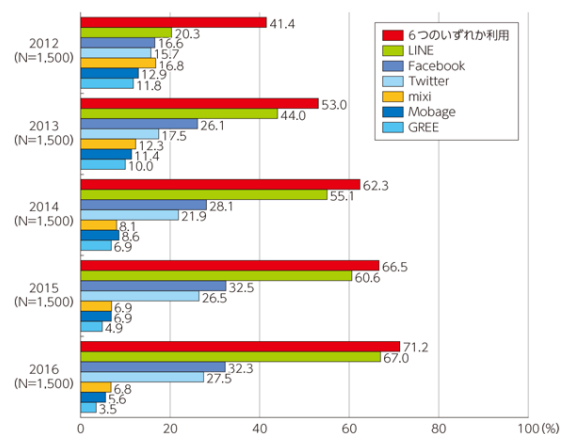


図 1 SNS 利用者数の推移[1]

身の見解を投稿できる機能がある。様々な価値観を持ったユーザが日常生活で感じたことについてユーザそれぞれの見解を投稿している。そのため，製品，サービスに対する口コミやレビューといった大衆の意見を得ることで製品やサービスの改善に役立てるための手段として利用されてい

1 大阪工業大学
Osaka Institute of Technology
2 立命館大学

Ritsumeikan University

る。製品、サービスに対する口コミやレビューといった大衆の意見を収集する手段として、図2意見収集手段の割合[6]のデータがある。意見を収集する手段として約7割が検索サイトあるいはSNSを利用している。

また、製品やサービスだけでなく、企業や行政が推進する何らかの取り組みに対する関心の把握にも利用されている。このような製品、サービス、取り組みに対する消費者やそれらに意見を持ったユーザの見解を分析する手段として感情分析という手法が用いられている。感情分析（センチメント分析）とは、テキストを構成している単語や文節で分類し、それらが共通して出現する傾向、時系列などから有用な情報を得るテキストマイニングの一つである。そのテキストにおける記述者の意見が、ポジティブな意見であるか、ネガティブな意見であるか、そのどちらでもないニュートラルな意見であるかを分析する手法である。感情分析を用いた SNS 利用における炎上を抑制するためのシステムの構築を行う研究[7]が行われており、この研究では、ユーザの SNS 利用における投稿によって炎上するテキストには第三者から見て不快に感じる特徴的な言語表現が多く存在し、それらの特徴的な言語表現を特定することで炎上の原因となるテキストの訂正を行えることが分かった。この研究から、感情分析は SNS におけるユーザの投稿から感情を推定できる技術であることがわかる。感情分析を用いてユーザの関心の変化が分かれば製品やサービス、取り組みの改善に繋げることができる。

本研究では、感情分析を用いることで Twitter のテキストから製品やサービス、取り組みに対する関心の変化の推定を行う。また、その関心の変化の手がかりとなる表現を抽出することでその要因の推定を行う。

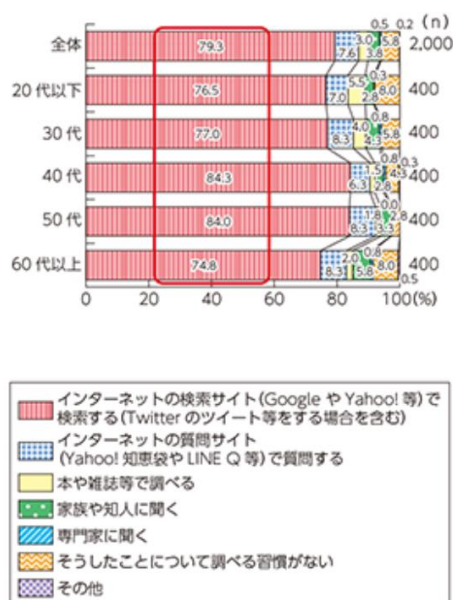


図2 意見収集手段の割合[6]

2. 関連研究

2.1 テキストに含まれる感情に関する研究

テキストにおける発話者、書き手の感情の推測がテキストの内容理解に繋がるが、テキストを読む過程で受け手には発話者がテキストに含めた感情とは異なる感情を認識するという研究がある。衣川ら[8]の研究では、日本語の文章読解授業において、テキストから印象に残った言葉をキーワードとして選び、学習者間でそのキーワードの背景にある登場人物の感情や意図に対する認識を調査した。その結果、読み手の中で価値観の創造が促されることを示した。

また、米田ら[9]の研究では、テキストの理解において読者の感情が、違和感・予感とは物語の展開の理解に繋がり、共感とは登場人物の心境の理解に繋がることを明らかにした。これにより、違和感や予感とは、一文ごとに情報を理解しようとする際に感じ、共感とは、テキスト全体に対して理解を示した際に感じることを示した。

これらの研究結果により、テキストを理解するには感情を理解する必要があると考えられるため、感情分析を用いてテキストへの理解の促進を図ることは有効であると考えられる。

2.2 感情変化の分析に関する研究

感情の変化を捉えるためには、時系列における数値化を行い、その数値の動的変化を捉える必要がある。そこで、感情の変化を捉えるために特定対象における世間一般の評判を数値化することが有効である。岩崎ら[7]の研究では、日次極性 (daily polarity: dp) を提案している。これは、一日ごとのポジティブなツイート数とネガティブなツイート数の差をトピックに関連した全てのツイート数で割った値である。トピックとは、発言に含まれる話題のことであり、本研究においてはトレンドのことを示す。Twitter におけるトレンドとは、Twitter 独自に選出されたトレンドキーワードのことである。リアルタイムで更新されているため、時間経過とともに変化し、Twitter 上で今最も盛り上がっているキーワードがトレンドとされる。dp は、そのトピックの全体的盛り上がりの中で、当日のツイートから読み取れる世間の評判を極性の強さで表している。dp に対し、その取得範囲の最大値で割ることで正規化したものが ndp (normalized dp) である。

2.3 感情分析手法と利用に関する研究

感情分析手法とは文章が包括する感情極性 (望ましいか否か) を判定することである。これは、質的データである文章が感情を包括する感情極性値として数値化し量的データとして扱えるようにする技術である。感情分析手法とその利用例として日本語での感情分析手法を先行研究として引用する。山内ら[10]の研究では、感情極性辞書や形態素解析エンジンの一つである McCab (メカブ) [11]、また、それをもとに開発された係り受け言葉解析器エンジンの一つ

である Cabocha (カボチャ) [12]を使用した Twitter での感情分析とその結果を利用することでテレビ番組の索引システムを提案した。

使用された感情分析手法は、放送中番組に関連するツイートを取得し、そのツイートを MeCab にて意味を持つ言語の最小形態である形態素に分解した、そして形態素を感情極性辞書と照らし合わせることで値を求め、それぞれの値の合計値によってツイートの感情値が算出されている。

Cabocha を使用することで感情語周辺の係り受け関係をあらかじめ定義した係り受けルールにより、感情値に重みづけを行った。例えば、「とても楽しい」という文において、「とても」が「楽しい」を強調する副詞であるため、「楽しい」の感情極性値が大きくなるよう重み付けされる。また、「全く楽しくない」という文は「全く」と「ない」が「楽しい」に係る反対の意味となるため、「楽しい」の感情極性値は負の値になるように重み付けされた。

高村らの[13]研究では、単語の感情極性を判定するために、コーパス等によって構築した語彙ネットワークによりスピン系のモデルを作成する手法を提案した。この研究でのスピンとは複数の電子の持つ方向であり、隣り合った電子はエネルギー的に同じ方向を向きやすいとされることから、このスピンを単語の感情極性とし、近似確率分布関数を計算することで単語の感情極性を調べた。

その結果を単語感情極性対応表としてまとめた。これは、岩波書店の国語辞典をもとに作成されており、全 55,125 単語が登録されている。単語感情極性対応表は、単語名、ふりがな、品詞、感情極性値で構成されている。品詞は、動詞、名詞、形容詞、副詞の 4 種類に分類され、これらに分類されないものは感情極性値が割り当てられていない。単語の感情極性値は、単語毎に、-1 に近似するほどネガティブな表現、1 に近いほどポジティブな表現として定義される。

本研究では、Twitter トレンドを含むツイートの関心変化を推定するために感情推定を行うので、膨大な量の日本語表現を扱うことになり、一つ一つの単語に感情極性値を算出することは困難であるため、感情分析手法として国語辞典をもとに全 55,125 単語に極性値が割り当てられている高村ら[13]の研究で作成された単語感情極性辞書と MeCab による形態素解析とによって感情分析を行う。

3. 提案手法

本研究での処理フローを図 3 に示す。TwitterAPI を用いてデータを収集後、収集したデータに対し、感情分析を行う。その後、Twitter トレンドに対する関心の変化を分析し、Twitter トレンドに対する関心を変化させた要因となる言語表現の推定を行う。



図 3 本研究の処理フロー

3.1 ツイートデータ収集

2021 年 1 月 31 日から過去一週間における Twitter トレンドとなった単語を含むツイートに対して TwitterAPI を用いてデータを取得する。取得できるツイート数は最大探索時間が TwitterAPI の仕様上約 15 分であるため、任意の Twitter トレンドとなった単語を含むツイートの収集は約 15 分で終了とする。約 15 分未満の時間で収集できる場合は、過去一週間における任意の Twitter トレンドとなった単語を含むツイート 100 件を収集する。

また、他人のツイートをそのまま引用して投稿するリツイートに関してはデータ収集から除外している。収集データにリツイートを含めることによる探索時間の増加と複数の同一情報を別情報としてカウントすることを予防するためである。

Twitter トレンド機能は、アルゴリズムで決定されるようになっており、Twitter を使用しているユーザが興味を持っているもの、フォローしているユーザアカウント、ユーザの位置情報をもとに選出される。新規 Twitter アカウントを作成し、提供する情報を絞ることで、フォローユーザなし、既存のツイートへの興味を抱いておらず、位置情報は日本国内となり、日本国内全体において盛り上がっている話題が Twitter トレンドとして表示される。

また、数日間やその日一日で話題になったトピックではなく、今まさに注目されているトピックが選出されるため、Twitter 上で最新の話題が Twitter トレンドである。

任意の Twitter トレンドとしての選定基準は幅広い分野から選び、一週間においてそのトレンドに関するリツイートではなく、多くのツイートが投稿されているものとする。

任意の Twitter トレンドとして収集した単語を表 1 に示す。

表 1 任意に収集した Twitter トレンドの一覧表

Twitterトレンド	取得ツイート数	トレンド入り確認時刻
鉄腕ダッシュ	100	2021/1/30/20:00
進撃の巨人	100	2021/1/31/20:00
キスマイANNP	100	2021/1/29/14:00
政権交代	100	2021/1/30/20:00
緊急事態宣言延長	100	2021/1/29/14:00
ウルフルズ	100	2021/1/29/14:00

3.2 感情分析

対象のツイートを取得したのち、形態素解析を行い単語の単位に分解し、単語感情極性辞書により分解した形態素の感情極性値を算出し、その合計値を Twitter トレンドに対してツイート投稿者が感じていたことを肯定的であるか否定的であるかが分かるように極性値として数値化を行う。

例えば、Twitter トレンド「鉄腕ダッシュ」におけるツイートでは、「今週も鉄腕ダッシュを見逃してしまった」という投稿があったとする。このテキストは、「今週も/鉄腕/ダッシュ/を/見逃し/て/しまった」に分割でき、これら 9 個の形態素と単語感情極性辞書を照らし合わせることで、それぞれの形態素の感情極性値を算出し、その合計値をツイート「今週も鉄腕ダッシュを見逃してしまった」での Twitter トレンド「鉄腕ダッシュ」に対する感情とする。

3.3 関心変化分析

関心の変化を捉えるためには、時系列における数値化を行い、その数値の動的変化を捉える必要がある。そこで、岩崎ら[6]の研究で提案された日次極性を参考にツイート毎の極性を算出する。この極性は、ツイート毎のポジティブなツイート数とネガティブなツイート数の差を Twitter トレンドに関連した全てのツイート数で割った値である。また、そのツイートデータの取得範囲の極性の最大値で割ることで正規化し、取得データ間での感情極性値の変化に規則性を見出す。

データを離散化し、関心の変化を捉えることで、Twitter トレンドになった言葉が表す事象の状態や状況が変化している場合、例えば、あるテレビ番組に対する放送前に抱いている関心と放送後に対して抱く関心の変化を時系列データの差分を取ることで考慮する。

ツイート投稿者毎のバイアスによる違いを Twitter トレンドになった言葉が表す事象の状態や状況の変化として間違えないようにするため、離散化幅を 5 格子分割と大きく取った。

また、ツイート毎にツイートを構成する形態素の数の差が大きく、形態素の感情極性値の相加平均であるツイートに対する関心にツイートのテキストの長さ、形態素の数が影響しないように分析する形態素数を一定にすることで考慮した。

具体的には、各ツイートを構成する名詞、形容詞、動詞、副詞の順に 10 個の形態素の感情極性値の相加平均をツイートに対する関心の値とした。

3.4 関心変化の要因推定

Twitter トレンドを含むツイートに対して投稿者が含ませていた感情を Twitter トレンドである言葉への関心として捉える。その関心の変化に対して影響を与えているもの、つまり、各ツイートにおいて感情極性値を最も変化させている形態素である最も 1 に近似している極性値を取る単語、あるいは、その反対に最も 1 に近似している極性値を取る単語が Twitter トレンドである言葉への関心を変化させた要因として抽出する。

具体的には、5 つの離散値における最小値と最大値をとる際に感情極性値が 1、あるいは、-1 に発散する単語を要因表現として抽出する。

4. 実験

Twitter API を用いてトレンド入り確認後の 6 つの Twitter トレンドを含むツイートデータ 100 件を収集した。各 Twitter トレンドがトレンド入りしたことが確認できた時刻以降に収集を開始し、最も新しいツイートから時系列順に 100 件収集した。それらのツイートを一つのデータ長とし、5 分割したデータの始まりにあたるツイートを Twitter トレンドに対する極性値の変化を離散値として分析した。20 件ずつで分割し、1 分割目が最も新しく、5 分割目が最も古いデータとなっている。極性値の変化から一定の規則性を見出すために正規化を行った。

また、5 つの離散値における最小値と最大値をとる際に発散する単語を要因表現として抽出した。

5. 結果と考察

Twitter トレンドに対して感情分析の算出結果を表 2 に示す。数値が出なかった分析結果は「None」としている。

表 2 Twitter トレンドの感情極性値変化一覧表

Twitterトレンド	感情値の変化				
	1分割目	2分割目	3分割目	4分割目	5分割目
鉄腕ダッシュ	-0.5253	-0.1757	-0.5919	-0.5838	-0.1743
進撃の巨人	0.1227	0.3198	-0.0756	0.3876	-0.9606
キスマイANNP	-0.7301	-0.7457	-0.9484	-0.5611	None
政権交代	-0.4694	-0.7272	-0.6343	-0.7507	0.0402
緊急事態宣言延長	-0.6848	-0.3607	-0.5893	-0.8130	-0.4177
ウルフルズ	-0.5911	None	0.9635	-0.7640	-0.7824

各 Twitter トレンドを含むツイートに対する感情極性値をほとんどの分割データにおいて算出できているが、Twitter トレンド「キスマイ ANNP」と「ウルフルズ」においては感情値を算出できないという結果であった。これは、

ツイート内において画像や動画を含むツイートあるいは顔文字や日本語以外の言語表現を含むものであったため、極性値が算出できなかったと考えられる。

この算出結果を Twitter トレンドに対する関心の変化として図 4 に示す。

図 3, 図 4 より, 連続する分割データ間における極性値の差は, -1 から 1 の差分において Twitter トレンド「進撃の巨人」と「ウルフルズ」のように大きいと 1.4482~1.7275 であった。大きく変化しているものは関心の変化がわかる程の要因が Twitter トレンドである事象にあったと考えられる。これは, Twitter トレンド「進撃の巨人」では, その言葉が表す作品の登場人物へのポジティブな関心, ネガティブな作品内容への関心を含むツイートがあったため感情極性値として大きな差があったと考える。Twitter トレンド「ウルフルズ」では, その言葉が表す作曲家の曲へのポジティブな関心を含むツイート, 「ウルフルズ」というハッシュタグを含むがツイートユーザの生活でのネガティブな感情を投稿したツイートがあったため感情極性値として大きな差があったと考える。

Twitter トレンドを含むツイートに対して, ツイート投稿者が抱く関心を感情極性値として算出した結果ネガティブな感情であるものとして分類される傾向があった。これは, 単語感情極性対応表に登録されている単語には, ネガティブな感情極性値であるものがポジティブな感情極性値であるものより多いため, -1 に近似する極性値を Twitter トレンドの極性値の相加平均に含めるからだと考える。

Twitter トレンドである言葉への関心としてその変化に対して影響を与えている, 各データ間において感情極性値を最も変化させている形態素を算出した結果を表 3 に示す。表 3 より, 各トレンドにおける関心の変化として最も影響を与えている言語表現がわかる。しかし, Twitter トレンド「鉄腕ダッシュ」, 「進撃の巨人」において, ポジティブな要因表現として抽出した結果, 「鉄腕」, 「巨人」を抽出した。これは, テキスト内に登場する Twitter トレンドである固有名詞を形態素解析にかけているためと考えられる。また, 表 2 と表 3 より, ウルフルズの 3,5 分割目のツイート内において「笑」, 「待つ」のみの形態素にしか極性値がなく, ツイート内においてこれらの単語の主語が単語感情極性対応表にはないため, 極性値がない形態素として算出しているからだと考えられる。

Twitter トレンドにおける離散化したツイート毎の極性と極性の正規化を行った結果を表 4 に示す。

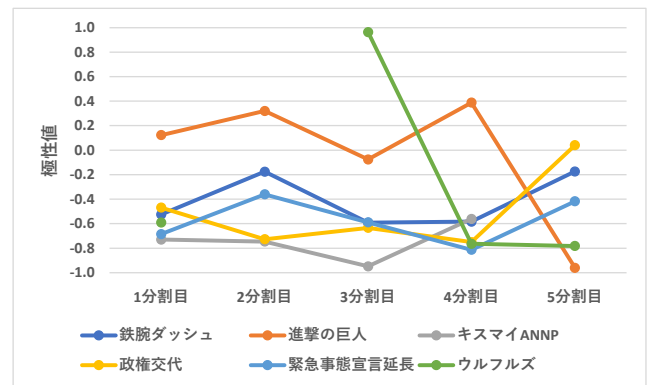


図 4 Twitter トレンドに対する関心の変化

表 3 Twitter トレンドの関心変化の要因表現

Twitterトレンド	ポジティブな要因表現	極性値	ネガティブな要因表現	極性値
鉄腕ダッシュ	鉄腕	0.9720	見逃す	-0.9868
進撃の巨人	巨人	0.9837	死ぬ	-0.9999
キスマイANNP	入所	-0.5610	心配	-0.9993
政権交代	好き	0.9320	変える	-0.9623
緊急事態宣言延長	調整	-0.2539	思う	-0.9023
ウルフルズ	笑	0.9635	待つ	-0.7824

ツイート毎の極性とその極性を正規化した値が大きいほど離散化したツイート毎に Twitter トレンドに対する関心に変化していると考えられる。ただし, ツイート毎の極性とその極性を正規化した値の差が大きいほど離散化したツイート毎と連続したツイート毎に離散化誤差が発生していると考えられる。

また, 100 件のツイートデータを離散化したことで Twitter トレンドである言葉が表す事象の状態, 状況が変化したことによる極性の変化であるかは現実での事象の状態とツイートの投稿時刻を確認する必要がある。本研究においてツイート収集時に Twitter API の制限上ツイート探索時間を短縮し, ツイート投稿時刻は収集できなかったため, 現実での事象の状態とツイートの関係性を確認することはできなかった。

表 4 ツイート毎の極性の正規化

Twitterトレンド	ツイート毎極性	極性の正規化
鉄腕ダッシュ	-0.1711	-0.4112
進撃の巨人	0.1902	0.4107
キスマイANNP	-0.0052	-0.0259
政権交代	-0.2005	-0.2535
緊急事態宣言延長	-0.0932	-0.2876
ウルフルズ	0.1629	-0.0943

6. おわりに

本研究では、SNSにおける関心の変化とその変化の要因となる言語表現の推定を行った。SNSにおける関心の変化としてTwitterトレンドに対する関心の変化を時系列データの離散値として定量化した。また、その変化の要因となる言語表現をTwitterトレンドへの関心である感情極性値に影響を与える、各ツイートにおいて感情極性値を最も変化させている形態素として抽出した。結果、Twitterトレンドを含む最も新しいツイートにおけるTwitterトレンドに対する関心の変化をリアルタイムの状況下において常に変化するツイート投稿者が抱いている感情の変化として捉えることができた。また、Twitterトレンドに対する関心変化の要因をTwitterトレンドに対する感情極性値の相加平均に最も影響を与えている形態素として抽出した。しかし、ツイート毎の投稿時刻と現実でのTwitterトレンドである言葉が表す事象の状況を考慮していなかったため、Twitterトレンドに対する関心の変化によって極性値に変化が現れたのか、ツイート投稿者毎のバイアスによるものか区別することはできなかった。今後の課題としては、データを離散的に捉えず、投稿後の時間経過が少ない少数の連続した時系列データからのTwitterトレンドに対する関心の変化を推定した場合、どのような条件下で行うべきかを検討する。また、ツイート投稿者がどのような感情を画像、動画に含ませているかの推定も必要である。

今後、他の情報媒体と本研究による結果を比較することでツイート投稿者毎のバイアスによる違いか、Twitterトレンドに対する関心の変化による違いであるかを検討する。

謝辞 本研究は、Twitter.Inc にツイートデータ収集の許可を得て行っている。

参考文献

- [1] “総務省 | 平成 29 年版 情報通信白書 | SNS がスマホ利用の中心に”, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/html/n_c111130.html, (参照 2020-12-20).
- [2] “Twitter”, <https://twitter.com/home>, (参照 2020-12-20).
- [3] “コミュニケーションアプリ LINE (ライン)”, <https://line.me/ja/>, (参照 2020-12-20).
- [4] “Facebook - ログインまたは登録”, <https://ja-jp.facebook.com/>, (参照 2020-12-20).
- [5] “Instagram”, <https://www.instagram.com/?hl=ja>, (参照 2020-12-20).
- [6] “総務省 | 平成 27 年版 情報通信白書 | 情報収集”, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/n_c122310.html, (閲覧日: 2021 年 1 月 30 日)
- [7] 岩崎祐貴, 折原良平, 清雄一, 中川博之, 田原康之, 大須賀昭彦: “CGM における炎上の同定とその応用”, 第 27 回人工知能学会全国大会, 2013.
- [8] 衣川隆生: 文章の登場人物の感情や意図を学習者はどのように解釈しているか, 日本語教育方法研究会誌, 2017.

- [9] 米田英嗣, 仁平義明, 楠見孝: 物語理解における読者の感情-予感, 共感, 違和感の役割, 日本心理学会論文誌, 2010.
- [10] 山内崇資, 林佑樹, 中野有紀子: 日本語解析による Twitter の感情分析とシーインデキシングへの応用, 情報処理学会第 75 回全国大会講演論文集, 2013.
- [11] “MeCab: Yet Another Part-of-Speech and Morphological Analyzer”, <http://taku910.github.io/mecab/>, (参照 2020-12-20).
- [12] “CaboCha/南瓜: Yet Another Japanese Dependency Structure Analyzer”, <https://taku910.github.io/cabocha/>, (参照 2020-12-20).
- [13] 高村大也, 乾孝司, 奥村学: スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, 2006.