

Twitterにおける解釈ニュアンスを含めた返信提示システムの提案

小山田健誠† 岡田佑一† 米澤朋子†
OyamadaKensei OkadaYuichi YonezawaTomoko

1. はじめに

近年、ソーシャル・ネットワーキング・サービス（以後 SNS）の利用者は増加傾向にある[1]。交流関係の維持や情報発信など様々な使用用途が存在し、個人利用者同士で社会情勢に対する議論が展開される場面もしばしば見受けられる。しかし SNS の一つである Twitter では、依然として攻撃的な投稿等による、特定個人または法人に対する批判・誹謗中傷が殺到する場面が存在する[2]。

このような Twitter における大規模な情報の拡散については様々な研究がなされており、小山ら[3]は炎上発生時にはユーザ同士で高密度のネットワークが形成されており、Twitter のユーザ間で同じトピックについてツイートが次々に伝播する共振構造を持つことを明らかにした。鳥海ら[4]は、炎上や災害等で特定的话题が大きく取り上げられるバースト現象時におけるトピックの分類とトピックについて情報を拡散しているコミュニティの分類によるバースト現象の詳細を明らかにする手法を提案した。

また Twitter で取得できるデータを用いた研究には、分析用データ取得だけでなく、様々な応用を前提とした文生成に関わるものがある。熊澤ら[5]は発話と応答のデータと発話者のプロフィールから学習データを作成し、深層学習による対話システムを提案した。岩倉ら[6]は雑話システムでの入力名詞に応じて、ツイートの集合内での単語間類似度からユーモアを生成するシステムを提案した。このように自然なコミュニケーションを演出する技術が Twitter のデータをもとに構築されている。

一方、短文での文章理解やそのニュアンスの変更に関する研究はまだ少ない。発話のニュアンスには複数の可能性があると考えると、固定の文章解釈のみを利用するシステムやコミュニケーション補助は、Twitter を通じたコミュニケーションの可能性の広がりや阻害するとも考えられる。そこで本研究では、単語の持つ印象の度合いを記した感情極性辞書[7]を用いて、3 つから 5 つ程度のツイートに対するニュアンスとして考えられるキーワードを選択肢として提示し、選択したキーワードのニュアンスを前提としたコミュニケーションを共有するシステムを提案し、その使用による理解度の向上や、対象ツイートに対する対話継続性の変化を検討する。

このようなキーワードの提示による文章理解については、キーワードの強調が強調内容の再生を高めるという研究結果がある[8]ことから、提案システムにおいても対象ツイートの文章理解を促せる可能性が考えられる。また提示したキーワードのバリエーションやその選択がユーモアとして受容されるのであれば、雑話システム[6]のような対話システムのように、ユーザに対話的に選択肢を与えその結果として他ユーザとのコミュニケーションを支援するシステムとなる可能性も考えられる。

2. 関連研究

2.1 文章理解と感情

文章理解について、文章における登場人物の感情の推測が文章の内容理解を促すこと、文章を理解する過程において読者の感情が異なることが報告されている。下記に詳細を述べる。

衣川[9]は日本語の読解授業において、学習者がそれぞれ文章から印象に残った言葉をキーワードとして選び、学習者どうしの対話を通してそのキーワードの背後にある登場人物の感情や意図を探る活動から、読み手の中に作品に関する意味世界を作ることが促されることを示した。

また米田ら[10]は物語理解にかかわる読者の感情について、違和感・予感とは物語の展開構造の理解にかかわり、共感とは登場人物の理解にかかわることを明らかにした。またこれに基づき、違和感や予感といった物語を進めるうえで起こる感情は、一文ごとの情報の理解の際に働き、共感といった物語を読み終えたか文章を理解した結果起こる感情は、文章全体の構造を理解した際に働くことを示した。

これらの研究結果により、文章理解の過程において感情が密接にかかわっていると考えられるため、感情分析を用いて文章理解の促進を図ることは有効であると考えられる。

2.2 感情分析手法

感情分析とは文章の持つ感情極性（望ましいか否か）を判定することであり、日本語の感情分析や性格分析についていくつかの手法を先行研究として引用する。

山内ら[11]は感情極性辞書や形態素解析器である MeCab^{*1}、及びそれをもとにした係り受け解析器の Cabocha^{*2}を使用した、Twitter での感情分析とその結果を利用したテレビ番組のインデキシングを行うシステムを提案した。ここで使用された感情分析手法として、まず放送中の番組に関連するツイートを取得し、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する、そして形態素を感情極性辞書と照らし合わせて得点を求め、それぞれの得点の合計値によってツイートの感情値を推定している、さらに Cabocha にて感情語周辺の係り受け関係を解析したのち、あらかじめ定義した係り受けルールに基づき感情値に重みづけを行っている、ここでの重みづけには、例えば、「とても楽しい」という文は、「とても」が「楽しい」に係ることで強調を表す副詞であるため、「楽しい」の感情極性値が大きくなるように重み付けされる。一方、「全く楽しくない」という文は「全く」と「ない」が「楽しい」に係り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る。

高村ら[12]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

†関西大学, Kansai University

*1 <https://taku910.github.io/mecab/>

*2 <https://taku910.github.io/cabocha/>

ピンを単語の感情極性とみなし、近似確率分布関数を計算し単語の感情極性を調べた。

塚野ら[13]は、Twitter 上の発言からニューラル言語モデルを利用した性格分析を行った。ここでは Twitter での発言と各ユーザを 5 因子モデルの性格指標で 5 段階に評価した。そして発言内容に対して MeCab を用いた形態素解析を行い、単語を主要 10 種類の品詞ごとに分類し、その各単語に対して word2vec^{*3} を利用して算出した 100 次元のベクトルを学習データとしてニューラルネットワークモデルに学習させた。

本研究では、後述する逆引き検索という手法をとるため、感情分析手法として山内らの研究[11]を参考とし、MeCab による形態素解析と感情極性辞書による照らし合わせを行う。

2.3 文章理解とキーワード

魚崎ら[14]は、文章の下線強調が、文章の難易度や文章の読解時間の長さにかかわらず文章の下線強調が強調部分の再生を高めることを示した。またこの研究で行われた実験では強調部分の短期再生において、あらかじめ文章のキーワードに下線が引かれている被験者群（プロンプト群）は読者が独自に文章に下線を引いた被験者群（アンダーライン群）より有効であった。このことからシステムによる概念提示は利用者の文章理解を促進する働きがあると期待できる。

砂山ら[15]は文章内で文章全体を特徴づけるキーワードを抽出するシステムを提案した。

また概念をあらわすシステムの提案に形容詞が用いられていることがあり、基本的概念をあらわす形容詞には特徴があることが示されている。

澤勢ら[16]は画像の周辺テキストの抽象度によって収集する画像の構造化を行うシステムを提案した。このシステムではユーザが画像を直感的に把握しやすいよう、画像と形容詞を対応付けるような識別器が設けられている。

海老原ら[17]は、もの、人、時間、空間、といった基本的な概念のカテゴリーを修飾する形容詞について、概念カテゴリーの抽象性の高さによってそれを修飾する形容詞の理解度に差が出ることを示した。

以上の研究から、文章から抽出するキーワードは文章理解を促進すると考えられる。また概念や状態の把握に形容詞を用いることは有効であると考えられる。

3. 提案手法

3.1 本研究の課題と手法の概要

本研究の目的はツイートのニュアンスを抽出、提示することによって、ツイートに対する印象や対話継続性がどのように変化するか検証することである。

この目的の達成には、発言内容から考えうるニュアンスを、システムのユーザに選択肢として提示する必要がある。ここには、提示する選択肢の量に関する以下に述べるような課題がある。

まず既存の対話システムのようなひとつの発言に対して一つの解釈しかないとする場合、発言に対して各々が別々の解釈を考える対話の多様性やユーモアが失われ、一方的な議論になる可能性がある。しかし一方で、発言に対し考えうる全てのニュアンスを提示した場合、多すぎる選択肢の中からユーザの視点を選択することが難しくなり、シス

テムのユーザビリティの低下や対話の発展に寄与しない可能性もある。

これらの観点から解釈ニュアンス提示システムには、発言内容に対して多過ぎずかつ一定に定まらない適度な数の解釈の選択肢が提示されることが求められる。

この課題に対して本研究では、関連研究にあるような感情分析を利用し、図 1 の様に感情極性辞書内の一定範囲にある 3~5 つの形容詞を提示する。これにより、発言内容にある程度添いつつ、ユーザが選択し返信しやすく、また、返信自体の多様性を生じやすい様になっている。

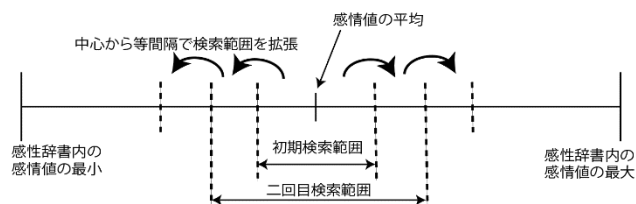


図1 逆引き検索範囲の変化

3.2 実装

この機能の実装を踏まえた、システムについて以下に詳細を説明する。ツイートを取得後、前章 2.2 項にて説明した感情語の感情極性辞書との照らし合わせによりツイートの感情値を決定する。そして感情極性辞書からその感情値に近い形容詞を複数取得し、ユーザに提示をする。この複数の形容詞から、ユーザが解釈の選択肢として一つ選択し、取得したツイートに返信するシステムを提案する。

まずシステムの具体的なフローとしては、

1. ユーザの Twitter アカウントでログインを行う
2. 対象のツイートを取得する
3. 形態素解析器 MeCab にて単語の単位に分解する
4. 単語を感情極性辞書に照らし合わせて 1 文ごとの感情値を取得する（感情分析）
5. 取得した感情値の平均に近い値を持つ形容詞を感情極性辞書から検索する（逆引き検索）
6. 取得した複数の形容詞をニュアンスとして提示する
7. ユーザの選択した一つの形容詞をリプライとして送信する

という一連の処理を通じて、解釈を選択する。このフローを図 2 に示す。

3.2.1 文章解析と選択肢の複数形容詞の選定

まず対象のツイートの取得について述べる。ユーザから希望のツイートの全文 URL を入力後、サーバに送信してもらうことで取得する。次に取得したツイート文を形態素解析器 MeCab を用いて、単語一つ一つに分解する。その後分解した単語と感情極性辞書内の感情の値を持つ単語（感情語）を照らし合わせ平均を得てツイートの感情値とし、その値に基づき感情極性辞書内より後述する逆引き検索に基づき条件に合う形容詞を抽出する。そしてそれらの形容詞を解釈の選択肢としてユーザに提示し、ユーザが選択した形容詞が対象ツイートの解釈として選択されたことがわかるような文章にユーザが返信文章を追記してツイートを送信する。

*3 <https://github.com/dav/word2vec>

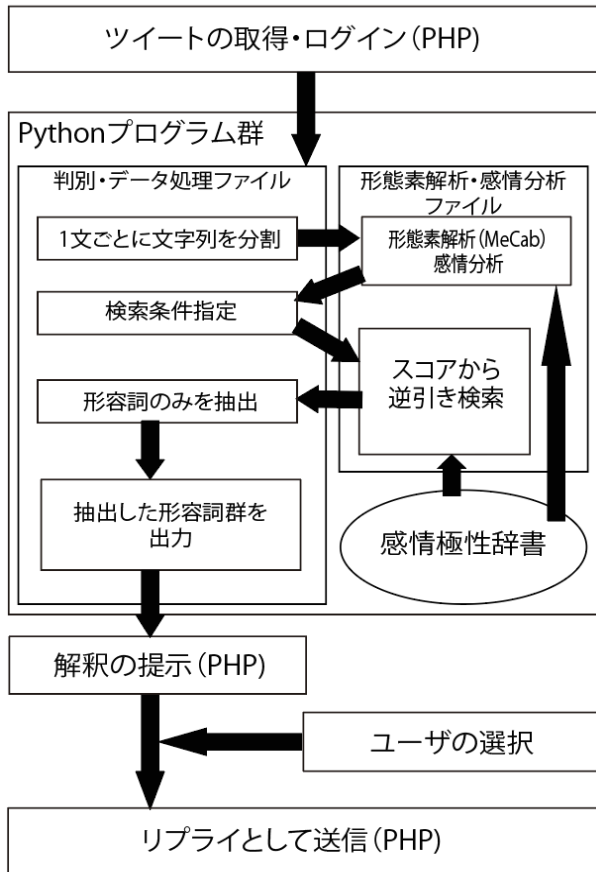


図2 システムフロー図

ここで、感情極性値から形容詞を見つける逆引き検索機能について説明する。ツイートの感情値に近い感情値を持つ形容詞が、ツイートのニュアンスとして候補になると仮定する。これに基づき、ツイートの感情値に近い値を持つ形容詞を3~5つ程感情極性辞書から検索、抽出を行う。

抽出する方法について図1に基づき説明する。あらかじめ細かい値で決めておいたツイートの感情値を中心とする探索範囲内で、形容詞の探索を行い、範囲内の形容詞が3つ未満の場合に、探索範囲をツイートの感情値を中心として細かい値で拡張する。この探索と拡張を繰り返す、範囲内の形容詞が3つ以上となった場合、それらすべての形容詞をユーザに提示する。

各種動作に使用する言語は、ツイートの取得、ユーザのシステム上 Twitter ログイン、リプライの送信を PHP で実装した。また、形態素解析、感情極性辞書に関する処理、を Python で実装した。また Twitter の API を利用して、システム利用者の Twitter アカウント情報を利用したツイートの取得、返信の投稿を行うため、システム利用者は事前に Twitter アカウントを持っていることが前提となる。

またこのシステムは Twitter 連携アプリケーション「ぶぶ漬け」としてサーバで処理を行うため、ユーザは各自の端末及びブラウザでのアクセスにて使用が可能になる。なお使用するサーバは GMO インターネット株式会社の提供するホスティングサービス「CONOHAVPS」を利用した。

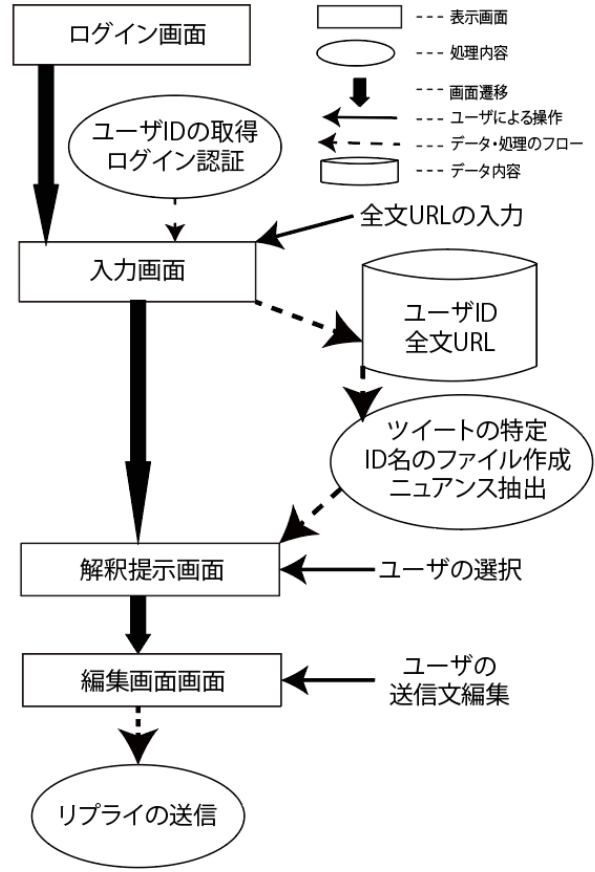


図3 インターフェースフロー図

3.2.2 ユーザインタフェースと実行結果

以下に実際に本システムを利用した際の入力、出力処理や画面遷移について説明する。このフローを図3に示す。

図4は本システムを利用する際に最初にアクセスする、ログイン画面の web ページである。本システムではシステム利用者の Twitter アカウントを利用したツイートの取得と返信を行うため、利用者のログインによって Twitter の API を用いた読み書き権限の取得を行う必要がある。したがってまず茶色の「Twitter にログイン」ボタンを押すことで、Twitter アカウントの認証を行い、そのユーザの Twitter ユーザ ID を取得したのち、図5に示す入力画面の web ページへ遷移する。

この画面でシステム利用者が入力する URL の入力に関して、Twitter の全文 URL の末尾の数字列はツイート固有の ID であることを利用して対象ツイートの特定を行うため、短縮 URL ではなく全文 URL を入力する必要がある。

入力画面の下部の空欄に URL を入力し送信ボタンを押すことで、サーバにシステム利用者の Twitter ユーザ ID、返信対象ツイートの URL が送信され、対象ツイートの特定とツイート全文の取得が行われる。

この際、取得した Twitter ユーザ ID の名前のテキストファイルをサーバ内で作成し、ツイート全文の文字列を記入する。このファイルを指定して、Python によるデータ処理プログラムを動作させることで、システム利用者ごとに動作結果を変化させている。

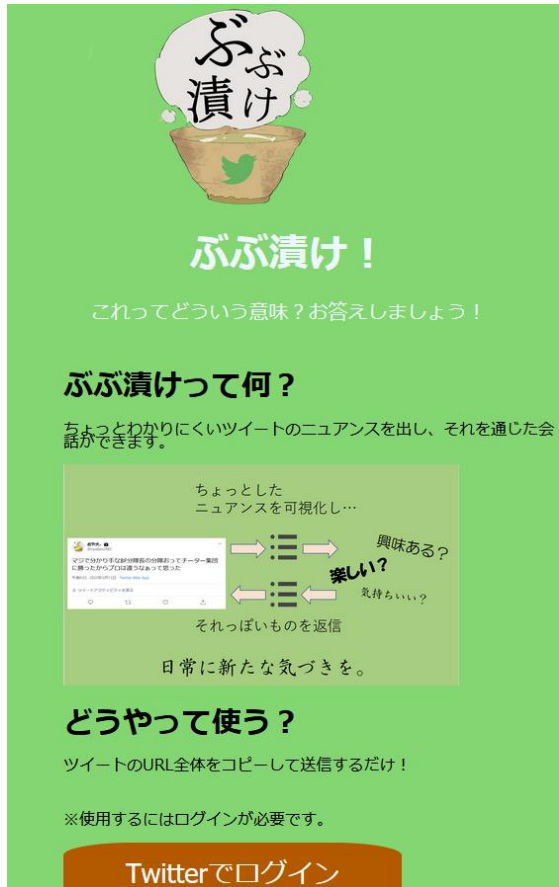


図4 ログイン画面

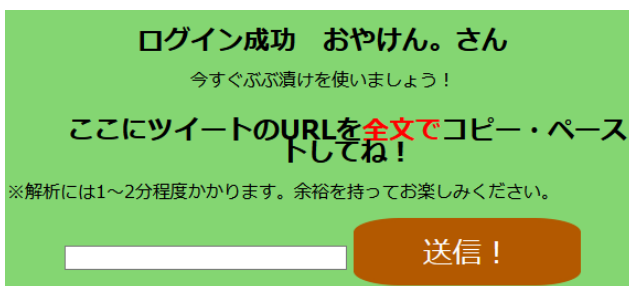


図5 入力画面

そして解釈ニュアンスの抽出が終了した後、図6に示す選択肢の提示画面に遷移する。この画面で対象ツイートの確認と選択肢の選択を行った後、送信ボタンを押すことで図7に示す送信ツイート編集画面に遷移する。編集画面の下部の入力欄には前画面で選択した解釈ニュアンスがあらかじめ入力されており、これをもとに返信する内容を編集できる。返信内容を編集し送信ボタンを押すことで、対象ツイートへのシステム利用者のアカウントでの返信がなされる(図8)。

これらの画面はPHPファイル内のHTMLの記述によって表示され、送信ボタンの色の指定などは対応するCSSファイルの記述によって変化させている。



図6 解釈提示画面

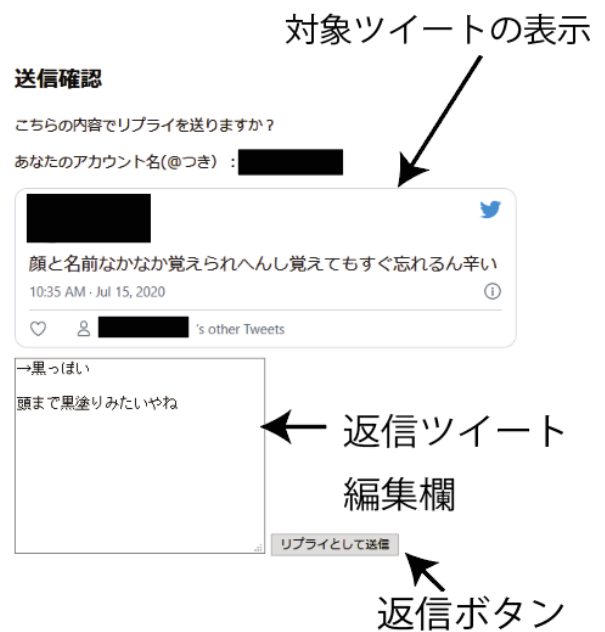


図7 編集画面

3.3 課題

これらのシステムの動作や処理での課題は、感情極性辞書の調整や提示する選択肢の変化が挙げられる。まず感情極性辞書について、本システムでは小林らの作成した感情極性辞書[19]をそのまま利用しているため、Twitterでの言語表現に則していない可能性がある。これにより感情分析の精度が失われ、適切な解釈ニュアンスの提示を行えない可能性がある。これに対し、Twitterの投稿データというドメインに基づいた感情極性辞書の拡張[18]によって対処できると考えられる。

また、提示する選択肢について、現在は感情極性辞書内の取得した感情値を中心とする範囲に存在する形容詞を取

り出しているが、本研究の目的はニュアンスの提示による解釈の多様化や対話続性の変化にあるため、取得した感情値とは異なる形容詞も選択肢に取り入れることを検討すべきであると考えられる。例えば「気に食わない」というネガティブなツイートに対して「大好き」といった極端にポジティブな解釈ニュアンスを選択肢として挙げた場合、「汚い」「暗い」といったネガティブな選択肢を選択して会話を行うよりも、意外性を持った会話の発展に寄与できる可能性がある。



図8 返信例

4. おわりに

本研究ではツイートの感情分析を行うことによって、一つのツイートに対する複数のニュアンスを提示し返信できるシステムを提案した。感情分析時に使用した感情極性辞書から条件に応じた形容詞を抽出し提示することで、複数の選択肢による多義性を持たせた会話が可能となっている。

今後の課題として、感情分析に用いる感情極性辞書には対象ドメインにてよく用いられる語を収録した感情極性辞書の収録が必要になる[18]ことから、Twitterに適した感情極性辞書を作成する必要がある。また文章全体での感情分析の精度向上のために、係り受け解析ルール[11]のような文構造によって感情値に補正を行うことが必要である。更に、ネガティブなやりとりの中に可能性を広げ対話継続性を高めていくためには、提示する形容詞をあえて分析によって得た感情値から離れた値を持つものにする(例えば、文章の感情値が-0.95であった場合、0.85に近い値を持つ形容詞を探索する)ことによって、より多くの意味が考えられる対話システムになる可能性が考えられる。

謝辞

本研究は一部科研費 19H04154, 18K11383, 19K12090, および 25700021 の助成を受け実施した。

参考文献

- [1] 総務省:平成 30 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(online), 入手先<https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000082.html>(2019. 9. 13)
- [2] 吉野ヒロ子, 小山晋一, 高田倫子:ネット「炎上」における情報・感情拡散の特徴:Twitter への投稿データの内容分析から, 広報研究学会誌委員会編(2018. 3)
- [3] 小山耕平, 浅谷公威, 榊剛史, 坂田一郎:ネット炎上におけるユーザーの共振構造, 人工知能学会全国大会論文集(2019. 6. 1)
- [4] 鳥海不二夫, 榊剛史:バースト現象におけるトピック, 情報処理学会論文誌(2017. 6. 15)
- [5] 熊澤優大, 大塚哲平, 谷津元樹, 原田実:対話者プロフィールと背景モデルを用いた深層学習による対話システム DeepEVE II, 情報処理学会第 81 回全国大会講演論文集(2019. 2. 28)
- [6] 岩倉亮介, 吉川大弘, フェリックスジメネス, 古橋武:雑話システムにおけるユーモア生成の受容性向上に関する検討, 情報処理学会第 79 回全国大会講演論文集(2017. 3. 6)
- [7] 森田晋也, 白井靖人:分野別単語感情極性辞書の作成及び評価, 情報処理学会第 80 回全国大会講演論文集(2018. 3. 13)
- [8] 関友作:テキストの内容把握に対する簡条書きとキーワード強調の影響, 日本教育工学会誌(2017. 10. 20)
- [9] 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. 5. 18)
- [10] 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, 心理学研究(2010. 7. 16)-
- [11] 山内崇資, 林佑樹, 中野有紀子:日本語解析による Twitter の感情分析とシーン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. 3. 6)
- [12] 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. 2. 15)
- [13] 塚野駿, 柴田千尋, 政倉祐子, 田胡和哉:ニューラルネット言語モデルによる Twitter 上の発言からの 5 因子モデルに基づく性格分析, 情報処理学会第 78 回全国大会講演論文集(2016. 3. 10)
- [14] 魚崎祐子, 伊藤秀子, 野嶋栄一郎:テキストへの下線ひき行為が内容把握に及ぼす影響, 日本教育工学会誌(2017. 10. 20)
- [15] 砂山渡, 谷内田正彦:文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌(2001. 2)
- [16] 澤勢一史, 延原肇:周辺テキストおよび感性情報に基づく形式概念分析を用いた画像群の概念構造化, 日本知能情報ファジィ学会第 28 回ファジィシステムシンポジウム講演論文集(2013. 7. 25)
- [17] 海老原聡子, 行場次朗:主要な概念カテゴリーを修飾する形容詞の感性特性, 日本認知心理学会第 12 回大会発表論文集(2014. 10. 5)
- [18] 小椋翔太, 桂井麻里衣:ニコニコ動画におけるスラングの感情極性辞書構築, 電子情報通信学会データ工学研究会 DEIM2017 論文集(2017)

- [19] 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(2011. 3. 1)