

令和元年6月12日現在

機関番号：16101

研究種目：若手研究(B)

研究期間：2015～2018

課題番号：15K16077

研究課題名（和文）不変情報と変動情報の融合による俗語感性辞書の構築

研究課題名（英文）Construction of a slang sensitivity dictionary by fusion of invariant information and variable information

研究代表者

松本 和幸（MATSUMOTO, Kazuyuki）

徳島大学・大学院社会産業理工学研究部（理工学域）・助教

研究者番号：90509754

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：本研究では、現在日本国内で使用されている日本語俗語を辞書などから収集し、マイクロブログ上の用例を収集することで大規模な俗語コーパスを構築した。このコーパスを用いて俗語の周辺文脈などからトピックモデルを作成し、俗語の用法のトピック変化や感性（否定/肯定など）の変化を分析することにより俗語感性辞書の自動構築をおこなった。また、俗語のなかでもネットスラングと呼ばれる語に関して、20代の被験者に対して複数の感性軸を対象とした感性評価アンケートを実施した。この結果をもとに、感性的な要素を考慮することで、意味要素のみを考慮する場合よりも意図をより正確に反映できる俗語の標準語変換手法を提案した。

研究成果の学術的意義や社会的意義

本研究結果により、従来は形態素解析の辞書など自然言語処理ツールでは解析対象とされにくかった俗語に対して感性と意味の両方の側面から分析し、標準語との相互変換が可能になった。この手法を用いることにより、たとえば、インターネット上の多種多様で膨大な意見集合から自動要約技術などにより意見を集約する際に、俗語を用いた発言の意図を意味と感性の両方を考慮して適切に解釈することが可能となる。また、俗語は、時間経過とともに新たな意味が付与されるため、本研究結果であるSNS上の発言からトピック変化や意味、感性の変化を分析する手法により、俗語辞書を構築すれば国語資料としても有用となることが期待される。

研究成果の概要（英文）：In this study, we collected Japanese slang words currently used in Japan from the dictionaries etc., and constructed a large-scale slang corpus by collecting examples on microblogs such as Twitter. By using this corpus, we created a topic model from the peripheral context of slang words, etc., and a slang sensitivity dictionary automatically by analyzing changes in topics and sensibility (negative / positive etc.) of usage of slang words. In addition, with regard to words called Internet slang, we conducted a sensitivity evaluation questionnaire for subjects in their 20s targeting multiple sensitivity axes. Based on this result, we proposed a method converting slang into standard word that can reflect intention more accurately than in the case of considering only semantic elements by considering sensuous elements.

研究分野：感性情報処理

キーワード：日本語俗語 自然言語処理 トピック分析 感性分析 ソーシャルメディア コーパス 標準語変換 時系列分析 自

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

自然言語処理技術を応用した様々なシステムが開発・実用化されてきているが、いまだ未知語が解析精度に及ぼす影響は大きい。とくに、近年流行している SNS における投稿には、辞書に登録されないような新しい俗語が数多く用いられ、かつ、口語で書かれた短文(ショートテキスト)が多く、従来の技術では新鮮かつ重要な情報を見逃してしまう可能性がある。インターネット上でよく用いられる俗語「リア充」や「イケメン」などのように、ひとたび一般ユーザにまで浸透してしまうと、形態素解析等の自然言語解析用の辞書への登録が必要になる。テキスト内の未知語の抽出や認識に関しては、これまでも研究されてきているが、俗語のみに着目した研究はほとんどない。俗語の多くが固有の物事を指し示すものではなく、既知語の組み合わせや類義の既知語への置換が可能な表現であるため、感性の認識において重要な手掛かりとなり得る。一方で、俗語のなかには、日々作成されては使い捨てられていく表現も多数存在し、すべてを辞書に登録していくのでは非効率的である。本研究では、あるコミュニティ、年代のユーザに、ある期間において一定以上の頻度で使用されている俗語を自動的に発見し、有用な情報を付与することで、評判分析や感情認識に応用可能な俗語感性辞書の自動構築技術を開発する。SNS 上で個人の発信した意見や情報が、発信者の意図に関わらず、第三者により拡散されてしまい、炎上という現象が起こることがある。これによって個人や団体の悪評が広まるのが新たな問題となっている。炎上の可能性のある発言に俗語が含まれていても、それを検出し対処する機能を実現するために、最新の情報が登録されている俗語感性辞書は有用である。

### 2. 研究の目的

最近の動向や世間一般の意見を探る対象として SNS 上の個人の発言は有用であるが、辞書に未登録の俗語が含まれることで解析に失敗し、必要な情報を見逃してしまう可能性がある。俗語の多くが感性を表現する言葉であるため、評判・意見分析では、俗語の感性辞書は重要である。また、新しく登場する俗語の多くは、使用されるコミュニティによって使用頻度や意味的な変遷があり、時間による変化が激しいため、感性とその用法や重要度をその都度変化させていく必要がある。本研究では、分析目的に応じて自動更新可能な俗語感性辞書の自動構築手法を提案する。本手法が実現することで、テキストマイニングのために毎回辞書を構築して準備する必要がなくなり、SNS における評判分析や炎上現象への対策などが従来より容易かつ正確におこなえるようになる。

### 3. 研究の方法

初年度から平成 28 年度までは、俗語辞書の自動構築に関する要素技術の確立をおこなう。ここでは、主に Twitter や動画コメントなどの Web 上のテキスト(変動情報)を自動収集するモジュールを構築し、収集したデータを用いて従来の俗語抽出手法を改良し、俗語スコア計算手法を確立させる。さらに、抽出した俗語に属性を付与するための知識ベースを構築する。具体的には、収集したテキストを語彙の重要度などを用いてフィルタリングしたあと、辞書知識と被験者によるアンケート調査の結果をもとに、俗語に感性情報を付与して俗語感性辞書を構築する。平成 29 年度以降は、作成した俗語感性辞書を用いて、未知俗語の感性判定や俗語の含まれる発言の炎上可能性を分析する手法について検討する。また、時間に余裕があれば、俗語以外のソーシャルメディア上の非言語情報に対しても解析手法を提案することで、より研究対象および応用先を広げることも検討する。

### 4. 研究成果

本研究により、主に以下の研究成果が得られた。

- (1) 日本語俗語とその印象や感性といった情報を登録した俗語感性辞書  
代表的な日本語俗語 671 語を選定し、複数の被験者(主に 20 代)を対象とした感性評価アンケートをおこなった。感性評価に用いた軸は、16 種類の感性評価対となっており、専用のアンケートツールを作成して調査した。
- (2) 構築した俗語辞書と標準語辞書をもとに俗語を標準語に変換する手法の確立  
(1) で構築した辞書と、日本語の既存の意味辞書(日本語語彙大系、分類語彙表、EDR 概念辞書、日本語 WordNet)から、俗語辞書に含まれる表現と意味が対応すると判断できる標準語を抽出し、俗語と標準語の対応関係を、感性類似度と意味類似度の 2 種類の類似度尺度を用いて判定することで、俗語から標準語に変換する手法を確立させた。
- (3) 俗語の変動解析用の時系列順に整理された大規模発話文コーパス  
俗語は時間の経過に伴い、使用文脈や意味用法が変化する。なかには、出現してからしばらくしてめったに使われなくなる、いわゆる「死語」や「廃語」といった語もある。これらの語の時系列的な用法変化を分析するために、アクティブユーザ数の多いソーシャルメディアの一つである Twitter を対象とし、Twitter API を用いて俗語を含んだツイート文の自動収集および、整理をおこなった。このデータに対するトピック分析や感情分析をお

こなうことで、俗語の意味用法、感性的な変化を検出することができた。

- (4) 未知俗語に対応するための表層特徴・文脈特徴に着目した俗語判別手法の確立  
辞書に登録されていない俗語に対しても感性や意味を分析できるようにするため、俗語の文字列特徴（文字の種類や並び）と文脈特徴（共起する単語の情報）を用いた手法を提案し、俗語かそれ以外の表現かを判定する手法を確立した。
- (5) 俗語と密接に関連する顔文字・絵文字・テキストアートの解析手法の提案  
俗語と同等かそれ以上にソーシャルメディア上で用いられる表現として、顔文字や絵文字などがあげられる。これらの表現を Web 上の非言語情報として意味や感性を解析することが可能であれば、俗語の分析にも役立つ。顔文字からの感情推定、絵文字やテキストアートからのカテゴリ推定などを深層学習をはじめとする機械学習によりおこなう手法を提案した。
- (6) 炎上検出のための未知表現に頑健な有害表現判定器の構築と有害表現辞書の構築  
ソーシャルメディア上の発言が発端となって生じる炎上現象を予測するため、発言内容に含まれる有害な表現（誹謗中傷や、差別表現など）を検出する手法を提案した。具体的には、人手により構築した有害表現辞書と大規模コーパスをもとに獲得した単語の分散意味表現を特徴量として、Support Vector Machine(SVM)を用いた分類器により有害表現判定器の構築を行った。

## 5 . 主な発表論文等

〔雑誌論文〕(計 14 件)

- [1] [Kazuyuki Matsumoto](#), Minoru Yoshida and Kenji Kita : “Emotion Recognition for Japanese Short Sentences Including Slangs Based on Bag of Concepts Feature Trained by Large Web Text,” Current Analysis on Instrumentation and Control, Vol.2019, No.2, pp. 9-18, 2019 (査読有).
- [2] [Kazuyuki Matsumoto](#), Fuji Ren, Masaya Matsuoka, Minoru Yoshida and Kenji Kita : “ Slang Feature Extraction by Analyzing Topic Change on Social Media,” CAAI Transactions on Intelligence Technology, Vol.4, No.1, pp.64-71, 2019 (査読有).
- [3] [Kazuyuki Matsumoto](#), Akira Fujisawa, Minoru Yoshida and Kenji Kita : “ ASCII Art Classification based on Deep Neural Networks Using Image Feature of Characters,” Journal of Software, Vol.13, No.10, pp. 559-572, 2018 (査読有).
- [4] Naoya Fujino, [Kazuyuki Matsumoto](#), Minoru Yoshida and Kenji Kita : “ Emotion Estimation Adapted to Gender of User Based on Deep Neural Networks,” International Journal of Advanced Intelligence (IJAI), Vol.10, No.1, pp. 121-133, 2018 (査読有).
- [5] Hirokimi Fukuda, [Kazuyuki Matsumoto](#), Minoru Yoshida and Kenji Kita : “ Index Generation of BGM Video Based on Distinctive Comments,” International Journal of Advanced Intelligence (IJAI), Vol.10, No.1, pp. 113-119, 2018 (査読有).
- [6] [Kazuyuki Matsumoto](#), Seiji Tsuchiya, Miyake Takeshi, Minoru Yoshida and Kenji Kita : “Flame Prediction Based on Harmful Expression Judgement Using Distributed Representation,” International Journal of Technology and Engineering Studies, Vol.4, No.1, pp. 7-15, 2018 (査読有).
- [7] [Kazuyuki Matsumoto](#), Fuji Ren, Minoru Yoshida, Kenji Kita, “ Review Score Estimation Based on Transfer Learning of Different Media Review Data,” International Journal of Advanced Intelligence, Vol.9, pp.541-555, 2018 (査読有).
- [8] [Kazuyuki Matsumoto](#), Hayato Shimizu, Minoru Yoshida, Kenji Kita: “ Time-Series Analysis of Video Comments on Social Media,” Sociolinguistics - Interdisciplinary Perspectives, Vol.8, pp.105-123, 2017 (査読有).
- [9] 松本和幸, 土屋誠司, 芋野美紗子, 吉田稔, 北研二: “ 感性を考慮した日本語俗語の標準語変換”, 人工知能学会論文誌, Vol.32, No.1, pp.1-12, 2017 (査読有).
- [10] [Kazuyuki Matsumoto](#), Satoshi Tanaka, Minoru Yoshida, Kenji Kita, Fuji Ren: “ Ego-state Estimation from Short Texts Based on Sentence Distributed Representation,” International Journal of Advanced Intelligence, Vol.9, No.2, pp.145-161, 2017 (査読有).
- [11] [Kazuyuki Matsumoto](#), Akira Fujisawa, Minoru Yoshida, Kenji Kita: “ Emotion Recognition of Emoticons Based on Character Embedding,” Journal of Software, Vol.12, No.11, pp. 849-859, 2017 (査読有).
- [12] [Kazuyuki Matsumoto](#), Fuji Ren, Minoru Yoshida and Kenji Kita: “ Refinement by Filtering Translation Candidates and Similarity Based Approach to Expand Emotion

Tagged Corpus,” Communications in Computer and Information Science, Vol.631, pp.260-280, 2017 (査読有).

- [13] Fuji Ren and Kazuyuki Matsumoto: “ Semi-automatic Creation of Youth Slang Corpus and Its Application to Affective Computing,” IEEE Transactions on Affective Computing, Vol.7, No.2, pp.1-16, 2016 (査読有).
- [14] Kazuyuki Matsumoto, Minoru Yoshida, Seiji Tsuchiya, Kenji Kita and Fuji Ren: “ Slang Analysis Based on Variant Information Extraction Focusing on the Time Series Topics,” International Journal of Advanced Intelligence, Vol.8, No.1, pp.84-98, 2016 (査読有).

〔学会発表〕(計 26 件)

- [1]. 藤野 尚也, 松本 和幸, 吉田 稔, 北 研二 : Word Mover’s Distance を用いたコーパス拡張による感情推定精度向上の検討, 言語処理学会第 25 回年次大会(NLP2019), 2019 年 3 月, 名古屋大学 (名古屋市, 愛知県).
- [2]. 羽田 優太, 松本 和幸, 吉田 稔, 北 研二 : リプライを用いたバズツイートの分類, 言語処理学会第 25 回年次大会(NLP2019), 2019 年 3 月, 名古屋大学 (名古屋市, 愛知県).
- [3]. Kazuyuki Matsumoto and Manabu Sasayama : Lyric Emotion Estimation Using Word Embedding Learned from Lyric Corpus, the 4th IEEE International Conference on Computer and Communications, Dec. 2018, (Chengdu, China).
- [4]. Akira Fujisawa, Kazuyuki Matsumoto, Ohta Kazuki, Minoru Yoshida and Kenji Kita : ASCII Art Category Classification based on Deep Convolutional Neural Networks, the 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nov. 2018, (Nanjing, China).
- [5]. Kazuyuki Matsumoto, Manabu Sasayama, Minoru Yoshida, Kenji Kita and Fuji Ren : Transfer Learning Based on Utterance Emotion Corpus for Lyric Emotion Estimation, 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nov. 2018, (Nanjing, China).
- [6]. 松本 和幸 : ソーシャルメディアの感性分析とその応用, 社会産業理工学研究交流会 2018, 2018 年 9 月, 徳島大学 (徳島市, 徳島県).
- [7]. 太田 万稀, 松本 和幸, 藤澤 日明, 吉田 稔, 北 研二 : 深層畳み込みニューラルネットワークによるアスキーアートの分類, 平成 30 年度電気関係学会四国支部連合大会, 2018 年 9 月, 愛媛大学 (松山市, 愛媛県).
- [8]. Kazuyuki Matsumoto, Minoru Yoshida and Kenji Kita : Classification of Emoji Categories from Tweet Based on Deep Neural Networks, the 2nd International Conference on Natural Language Processing and Information Retrieval (NLP-IR2018), Sep. 2018, (Bangkok, Thailand).
- [9]. 松本 和幸, 藤澤 日明, 吉田 稔, 北 研二 : アスキーアート分類手法の比較検討, 第 32 回人工知能学会全国大会, 2018 年 6 月 (鹿児島市, 鹿児島県).
- [10]. 松本 和幸, 任 福継, 吉田 稔, 北 研二 : 絵文字を手がかりとした俗語感情コーパスの拡張, 言語処理学会 第 24 回年次大会, 2018 年 3 月.
- [11]. 藤野 尚也, 松本 和幸, 吉田 稔, 北 研二 : ユーザの性別と感情表出傾向との関連, 第 31 回人工知能学会全国大会, 2017 年 5 月 (名古屋市, 愛知県).
- [12]. 藤澤 日明, 松本 和幸, 吉田 稔, 北 研二 : 画像特徴量を用いた大型アスキーアートの分類手法の提案, 第 31 回人工知能学会全国大会予稿集, 1-2, 2017 年 5 月 (名古屋市, 愛知県).
- [13]. 松本 和幸, 任 福継, 吉田 稔, 北 研二 : 感情表出傾向の分析による対話破綻検出, 言語処理学会第 23 回年次大会, 2017 年 3 月, 筑波大学 (つくば市, 茨城県).
- [14]. 三宅 剛史, 松本 和幸, 吉田 稔, 北 研二 : 分散表現を用いた有害表現判別に基づく炎上予測, 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(第 15 回), 2017 年 3 月, 東京大学 (東京都).
- [15]. Kazuyuki Matsumoto, Tanaka Satoshi, Minoru Yoshida, Kenji Kita and Fuji Ren : Ego-state Estimation from Short Texts Based on Sentence Distributed Representation, The 11th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE’16), Dec. 2016, (Okinawa, Japan).
- [16]. Kazuyuki Matsumoto, Minoru Yoshida, Kenji Kita, Yunong Wu and Iwasa Fumihiro : Effect of Users Attribute on Emotion Estimation from Twitter, the 2nd IEEE International Conference on Computer and Communications, Oct. 2016, (Chengdu, China).
- [17]. Kazuyuki Matsumoto, Seiji Tsuchiya, Minoru Yoshida and Kenji Kita : Judgment of Slang Based on Character Feature and Feature Expression Based on Slangs Context Feature, the 2nd International Conference on Soft Computing in Data Science 2016 (SCDS2016), Sep. 2016, (Kuala Lumpur, Malaysia).
- [18]. 田中 聡, 松本 和幸, 吉田 稔, 北 研二 : 情報推薦のための Twitter ユーザの性格分析手法, 人工知能学会全国大会 2016, 2016 年 6 月 (北九州市, 福岡県).
- [19]. 藤澤 日明, 松本 和幸, 吉田 稔, 北 研二 : 画像特徴量を用いたアスキーアートからの顔文字検出, 人工知能学会全国大会 2016, 2016 年 6 月 (北九州市, 福岡県).
- [20]. 岩朝 史展, 松本 和幸, 吉田 稔, 北 研二 : Twitter ユーザの属性別感情推定の検討, 言語

処理学会第 22 回年次大会, 2016 年 3 月 (仙台市, 宮城県).

- [21]. 松岡 雅也, 松本 和幸, 吉田 稔, 北 研二 : トピック変動の分析による俗語の特徴抽出, 情報処理学会研究報告, 2016 年 1 月, 国立情報学研究所, (東京都).
- [22]. Kazuyuki Matsumoto, Minoru Yoshida, Seiji Tsuchiya, Kenji Kita and Fuji Ren : Slang Analysis Based on Variant Information Extraction Focusing on the Time Series Topics, the 10th International Conference on Natural Language Processing and Knowledge Engineering, Oct. 2015, (Sapporo, Japan).
- [23]. Kazuyuki Matsumoto, Minoru Yoshida and Kenji Kita : Sensibility Estimation Method for Youth Slang by Using Sensibility Co-occurrence Feature Vector Obtained from Microblog, Proceedings of the 2015 IEEE International Conference on Computer and Communications (ICCC2015), Oct. 2015, (Chengdu, China).
- [24]. 岡本 拓馬, 松本 和幸, 吉田 稔, 北 研二 : 影響力の範囲を考慮した Twitter における影響力推定手法, 第 14 回情報科学技術フォーラム, 2015 年 9 月, 愛媛大学(愛媛県松山市).
- [25]. 前川 和輝, 松本 和幸, 吉田 稔, 北 研二 : 掲示板における炎上分析のためのノイズ除去手法, 第 14 回情報科学技術フォーラム, 2015 年 9 月, 愛媛大学(愛媛県松山市).
- [26]. Kazuyuki Matsumoto, Minoru Yoshida, Qingmei Xiao, Luo Xin and Kenji Kita : Emotion Recognition for Sentences with Unknown Expressions based on Semantic Similarity by Using Bag of Concepts, Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'15), 1428-1433, Aug. 2015, (Zhangjiajie, China).

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。