特集論文 「Web インテリジェンスとインタラクション 2017」

感性を考慮した日本語俗語の標準語変換

Conversion of Japanese Slang into Standard Japanese Considering Sensibility

松本 和幸 Kazuyuki Matsumoto
徳島大学 Tokushima University
matumoto@is.tokushima-u.ac.jp

土屋 誠司 Seiji Tsuchiya
同志社大学 Doushisha University
stsuchiy@mail.doshisha.ac.jp

芋野 美紗子 Misako Imono
大同大学 Daido University
m-imono@daido-it.ac.jp

吉田 稔 Minoru Yoshida
徳島大学 Tokushima University
mino@is.tokushima-u.ac.jp

北 研二 Kenji Kita
（同 上）
kita@is.tokushima-u.ac.jp

keywords: Internet slang, standard word, sensibility, concept of word

Summary

With the recent spread of communication using social media, exchanging opinions each other on web has become more common irrespective of age and sex. On the other hand, a problem called as “Internet flaming” often occurs along with the increase of social network service users. One of the reasons might be that the users do not recognize meanings/intentions/emotions expressed by other users’ words. In this study, we focused on slangs (Internet slangs) that are often used on SNS but are not registered in dictionaries, then tried to convert them into standard words. We also intended to output more appropriate candidates by considering not only semantic similarity but also affective similarity. The proposed method conducts filtering and re-ranking over the semantically similar candidates obtained based on distributed representations to detect the inappropriate candidates as standard word by focusing on two points: (1) features of slang/standard word and (2) affective similarity between the inputted word and the candidate words. In the evaluation experiment, the proposed method obtained a higher MRR than the baseline method.

1. は じ め に

近年のソーシャルメディアの普及に伴い，Web 上には，老若男女，職業等を問わず，多種多様な背景を持つ人物同士の意見交換が積極的におこなわれるようになった．こうした場における発言は，不特定多数の第三者が目にする可能性がある．そのため，ある発言が発端となり，発言者が批判の矛先となってしまう「ネット炎上」,「叩き」,「誹謗中傷」などのトラブルが相次いでいる．これらの言葉の暴力による問題は，Web 上のほうが起こりやすく，収まりにくい．原因のひとつに，Web 上の発言（言語情報）のみでは，発言者の意図や感情が伝わりにくいことが挙げられる．さらに，若者が多用する，一般に若者言葉と呼ばれる言葉は，主語，述語，語尾などを曖昧にしたり，省略語を用いることによる婉曲表現が多く，これらが相互の理解を妨げる要因になっていることは否めない．また，若者言葉のように，新たに作られた言葉は，辞書に登録されにくいため，すべての人が共通して理解できる言葉ではない．

辞書に登録されていない未知語の解析を目的とした様々なアプローチが，自然言語処理の分野において提案されている [鍛冶 13, 森 14]. しかし，これらの研究の目的は固有名詞（人名，施設名，製品名，etc.）や専門用語の処理が主であり，若者言葉のようにその特性上，明確に定義することが困難な語についての研究例は少なく，今後の進展が期待されている．

俗語の意味や感性的なニュアンスを理解できるか否かが，とくに若者が記述した文章からの意見・評判分析の精度に影響を与えると考える．また，若者言葉のなかには，標準語を言い換えた表現が数多く存在するため，若者言葉を標準語に変換（復元）することができれば，意見・評判分析や感情認識の精度改善に貢献できる．若者言葉は使われている期間や場面，コミュニティなどにばらつきがあり，意味・用法，感性的印象なども，時代とともに変化するものである．そのため，小規模な言語コーパスを分析するだけでは，その傾向を上手くつかむこと

ができない．したがって，とくに若者が発言することの多い Web 上のソーシャルネットワークサービス（SNS）から収集したテキストに基づきコーパスや辞書を動的に構築・更新できるような手法が重要となる．

本研究では，若者言葉を多次元の印象軸（感性評価軸）と，意味（概念）ベクトルによって表現することで，意味的にも感性的にも類似した標準語に変換することを目的とする．たとえば,「明日もバイト入ってて タヒる わ」という若者言葉を用いた文から，標準語を用いた文への言い換えを考える．意味のみを考慮した場合，標準語を用いた文は,「明日もバイト入ってて死ぬわ」といった表現となる．この例では，若者言葉の「タヒる」を「死ぬ」という直接的な表現に置き換えることにより，本来の意味の「死ぬ」といった，より深刻な状況を連想させてしまうため，感性を適切に表現できているとはいえない．このため,「明日もバイト入ってて しんどい わ」といった文に変換できることが望ましい．このように，意味のみではなく感性を考慮すれば，意味は多少異なるが感性的には近い変換候補の取りこぼしを減らすことを目指す．

以降，関連研究について述べ，若者言葉の印象（感性）について分析をおこなう．さらに，概念ベクトルの生成に用いる Tweet コーパスの構築と，提案手法である若者言葉から標準語への変換手法について述べ，ベースライン手法との比較による評価実験について説明する．最後に，結果について考察し，まとめる．

2. 関 連 研 究

若者言葉のみに限定した自然言語処理分野における研究は少ない [原田 02]．これは，若者言葉そのものの定義自体が明確なものではなく，言語学の分野での研究もあまり進んでいないことが原因であると考えられる．また，若者言葉を意味により分類する場合に，分類結果に対する正誤判定が困難であることも，研究対象としづらい原因と考えられる．しかし，辞書には登録されないような，擬音語・擬態語や，くだけた表現を処理する手法についての研究は従来から存在している [池田 10, 松尾 14, 三枝 07, 土屋 12, 内田 12]. これらの研究においては，深い意味解析はせずに，ルールやパターンマッチングによる処理で実現している．実際に，これらの手法を適用する場合には，ある程度限定された環境（文書のドメインなど）であることが前提となる．本研究では，対象となる文書のドメインの限定はおこなわないが，若者言葉が文中から抽出できているという前提で，標準語への変換を目的とする．また，本研究では，ルールやパターンマッチングを用いず，コーパスから学習した文脈特徴を単語の概念として用いる．

一方で，言語の感性的な印象に関する研究は近年，盛んになってきている．本研究では，若者言葉を標準語に変換する際，似た印象の言葉を優先して出力することを目指しているため，言語の感性的な印象を得る必要がある．

Matsumoto et al. [Matsumoto 11] は，若者言葉が表現する感情を推定するために，文脈情報および単語から得られる表層情報（文字種，単語の画数など）を用いた．

山西ら [山西 15] は，子供に付ける名前に対する感性的な印象（一般的/奇妙）を言語的特徴から判定している．この研究では，言語の表層的な特徴である「漢字の個数」,「読みの発音数」などを素性として用いて，Support Vector Machine による 2 値（一般的/奇妙）分類をおこなっている．

これらの研究に共通するものとして，単語から受ける印象などの感性情報を判断するために，あらかじめ単語に対する印象をアンケートにより取得することで正解データを作成していることがあげられる．本研究でも，若者言葉に対して抱く印象について事前にアンケートにより取得する方法をとる．

また，自然言語処理の分野において単語の言い換えの研究は従来からも多数おこなわれている [藤田 01, 野口 16]．藤田ら [藤田 01] は，語釈文を利用し，普通名詞を同概念語に言い換える手法を提案しているが，普通名詞を対象とし，同じ概念を持つ単語への言い換えが目的であるため，本研究のように辞書に載っていないことが多い俗語を対象とした研究とは目的が異なる．また，本研究では主に，ツイート文のようなくだけた口語文で書かれた用例文をもとに俗語の概念を学習させるが，藤田らの手法では，新聞記事コーパスに基づいて共起情報を得ることにより意味差分を獲得し，文脈における言い換えの可否を判定している点でも，本研究とは異なる．また，野口ら [野口 16] は，日本語複合動詞の言い換えを目的としており，本研究とは対象が異なる．

3. 若者言葉の感性分析

若者言葉は，仲間内の会話において，過激な内容の発言を柔らかい印象に変化させたり，言葉では表現しにくい状況などを伝える際に臨場感を持たせたりすることなどによく用いられる [米川 98]．一方で，標準語は，一般に，不特定多数の人に発話の意図や意味を正しく伝えることを第一目標としている．このため，意味的に同じか，あるいは類似する若者言葉と標準語が必ずしも同一の印象を与えるとは限らない．これは，表現を若者言葉に言い換えることで柔らかい印象を与えたり，物事を婉曲的に表現する作用を持っているためである．また，意外性や親密さを演出したり，照れ隠しをするためにも用いられることから，若者言葉は感性を豊富に表現できるものであるといえる．

本章では，それぞれの若者言葉がどのような印象を持っているか，また，それらの印象が標準語とどのような違いがあるかを，感性評価アンケートにより得られたデータに基づき分析する．得られたデータを，提案手法によ

図 1 アンケート回答ツールの GUI

表 1 アンケートに用いた感性評価軸

ネガティブ – ポジティブ	品が無い – 汚い
きたない – きれい	乱暴 – おだやか
暗い – 明るい	くだけた – 改まった
にくらしい – かわいらしい	軽薄 – 誠実
つまらない – 面白い	うるさい – 静かな
焦り – 安心	驚き – 期待
悲しみ – 喜び	恐れ – 怒り
鎮静 – 興奮	嫌い – 好き

り適した標準語候補が得られているかどうかを評価するために用いる．

3·1 若者言葉の感性評価アンケート

若者言葉が与える感性を，アンケートに対する複数の被験者の回答を分析することで調査する．本調査では，若者言葉感情コーパス [Matsumoto 12a, Ren 15] に含まれる若者言葉と，ニコニコ大百科 [ニコニコ大百科] において，語釈文に若者言葉，ネットスラング，隠語という表記のある見出し語を合わせて 671 語選定し，ランダムで 2 等分し，被検者 1 名あたり約 300 語について回答する形式とした．また，各語に対し少なくとも 2 名以上の被験者が回答するようにした．

アンケートの回答には専用の回答ツールを用いて，各表現に対し，16 種類の感性評価対を設け，各々50 段階で評価する．アンケート回答ツールの GUI 画面を図 1 に示す．また，表 1 に，アンケートに用いた感性評価軸を示す．

アンケート結果を 16 次元の感性評価ベクトルに変換･正規化し，各若者言葉間の印象の類似性を可視化するため，自己組織化マップ (SOM: Self-Organizing Map) [Kohonen 82] を用いた分析をおこない，一部の若者言葉を 2 次元座標平面上に配置したものを，図 2 に示す．この図をみると，よく似た印象の若者言葉が近い位置に表示されている．このことから，取得したアンケート結果がある程度信頼できることが分かる．

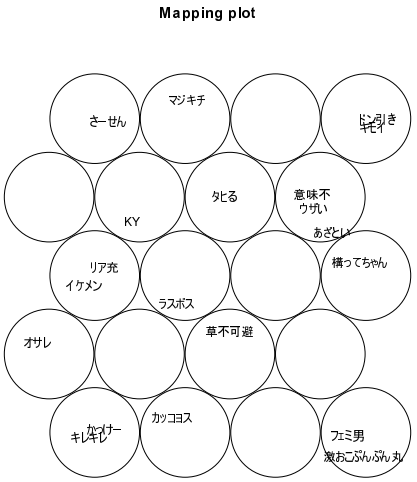


図 2 感性評価ベクトルに基づき自己組織化マップにより若者言葉を配置した例

3·2 若者言葉と標準語の感性比較

つぎに，若者言葉に対応する標準語との印象の比較分析をおこなうため，前述の感性評価アンケートにおいて用いた 671 語のなかから，以下の 2 つの条件に当てはまる若者言葉の抽出をおこなった．

- 同一表記の語（意味が異なるものも含む）が既存の標準語辞書には登録されていない
- 意味が同一または類似する表現が標準語辞書に登録されている

本研究では既存の標準語辞書として日本語 WordNet [Bond 12]，日本語語彙大系 [池原 97]，分類語彙表 [国立国語研究所 04]，EDR 概念辞書 [情報通信研究機構] の 4 つの辞書を用いた．抽出された語は 154 語となった．本節では，この 154 語の若者言葉に注目する．本来ならば標準語についても，若者言葉と同様の感性評価のアンケート分析をおこなう必要がある．しかし，同じ意味の標準語でも異なる表記で記述されることで異なる印象を与えることがあると考えられる．そのため，あらゆる表記を網羅したアンケートを実施することは現実的には不可能であることから，本研究では標準語に関しては positive/negative/neutral の感性（感情極性）のみを対象に比較分析をおこなう．標準語の感情極性が登録された言語資源として，高村ら [高村 06] の構築した感情極性値対応表や，乾ら [小林 05] の公開している評価表現辞書，佐野 [佐野 11] の構築した日本語アプレイザル辞書など，複数の有用な辞書が存在する．これらを相補的に用いて，今回の比較分析をおこなう．若者言葉と変換対象の標準語の positive/negative/neutral の組合せを集計した結果を図 3 に示す．横軸は若者言葉の感情極性，棒グラフは若者言葉と対応する標準語の positive/negative/neutral の内訳を示す．

この結果から若者言葉と標準語との感性が一致する割

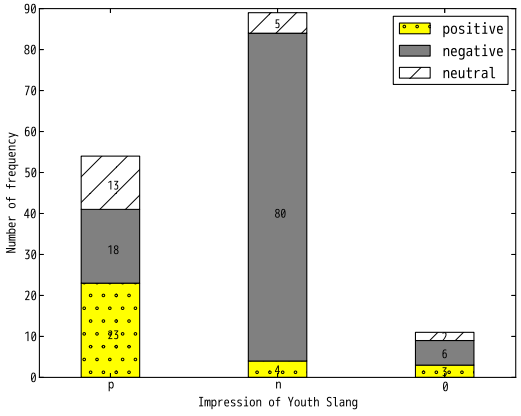


図 3 若者言葉と対応する標準語の感性（感情極性）の比較

合が 66.8%であることがわかった．一致しない場合もある程度みられることから，若者言葉から標準語へ変換することで感性が変化してしまう（positive/negative が反転する）可能性がある．感性が一致しない組合せにおいて，若者言葉が positive，標準語が negative の場合がもっとも多く，18 組あった．つまり，標準語では negative な意味にとらえられがちの語でも，若者言葉で表すことにより，positive な印象を与えることができるケースが多くあるといえる．

4. コーパス構築

4·1 若者言葉 Tweet コーパス

若者言葉は様々な場面で利用されるが，インターネット上におけるブログや SNS，電子掲示板などで多用される傾向にある．とくに，Twitter のように 1 つの投稿における文字数制限がある場合に，物事を端的に言い表せることから省略表現を含む若者言葉がよく使用される．たとえば,「現実世界で充実した生活を送っている人」を表す若者言葉は「リア充」であるが,「リアル」と「充実」の複合語を省略した構成となっている．

本節では，若者言葉の概念を用例から得るため，Twitter から自動収集したテキストデータにノイズフィルタリング処理を施し，若者言葉を含む発話文テキストを登録した若者言葉コーパスを構築する．若者言葉が含まれるか否かは，Web 上の若者言葉辞典 [若者言葉辞典] や，日本語俗語辞書 [日本語俗語辞書] から若者言葉として適切と思われる語を，その表記の違いなども考慮して作成したリストとの照合をおこなうことで判定する．このリストには，先行研究で構築されているコーパス [Ren 15] に登録される若者言葉に加えてランダム抽出により抽出した語を含めて計 1,323 語を登録している．自動収集の期間は，2014 年 12 月〜2015 年 6 月の 7 カ月間とした．構築したコーパスの基本統計情報を，表 2 に示す．本研究では，Twitter から収集したテキストデータの形態素解析を

表 2 若者言葉 Tweet コーパスの統計情報

発話文数	3,875,507
単語総数	86,932,177
若者言葉総数	11,290,873
1 発話文あたりの平均単語数	22.43
1 発話文あたりの平均若者言葉数	2.91

おこなうために MeCab ver.0.996 [MeCab] を用いた．若者言葉によっては，形態素解析によって正しく単語分割されない場合がある．たとえば,「ていうか」という若者言葉は,「て / いう / か」のように分割されてしまう．そのため，前述のリストに含まれている若者言葉に対して，テキストデータを形態素解析する前に若者言葉の前後に分割記号を挿入しておき，形態素解析後に，その分割記号に囲まれた文字列を一つの単語に連結するといった後処理をおこなった．

4·2 感情表現 Tweet コーパス

若者言葉をキーとして収集したコーパスは，発言しているユーザに偏りがある可能性が高い．その理由として，若者言葉のような特殊な語は，使う者を選ぶことがあげられる．また，本研究でのコーパス収集には Twitter のストリーミング API を用いており，複数の計算機での同時収集はおこなっていない．このため，ある単語が収集される時刻が偏ってしまい，多様な共起表現の収集がおこなえないことも考えられる．

本研究では，意味的のみならず感性的にも類似した変換候補を得ることを目的としているため，感性的な表現が多く含まれるであろうコーパスを別途準備する必要があると考えた．感性的な表現の代表的な語彙のリストを得るために，日本語アプレイザル評価表現辞書 [佐野 11] や感情表現辞典 [中村 93] に登録されている表現をピックアップし，これらをもとに，意味的な類似性のある語を分類語彙表 [国立国語研究所 04] と日本語 Web コーパス [Web Corpus] を組み合わせて拡張した感情表現リストの生成をおこなった．

分類語彙表では，同義または類義の語が得られる一方で，Web の口語表現ではあまり用いられないものも多数登録されているため，適切な拡張にならない場合もあると考えられる．そのため，分類語彙表によって得られた同義・類義語リストに対し，日本語 Web コーパスの N-gram から共起語を文脈ベクトルとして作成し構築した文脈類似語データをもとに，文脈類似語上位の語のみを残す処理を施すことで，拡張候補の絞り込みをおこなった．このようにして作成された感情表現リストに含まれる表現数は，15,322 となった．このリスト内の語を，一巡するたびにランダムに並べ替え，取得できる最大件数の Tweet を取得する．これを何度か繰り返すことにより，ある程度の量の Tweet を取得した．

また，取得した Tweet データから，ハッシュタグと判

表 3 感情表現 Tweet コーパスの統計情報

発話文数	5,291,498
単語総数	161,714,808
感情表現総数	6,163,139
1 発話文あたりの平均単語数	30.56
1 発話文あたりの平均感情表現数	1.589

別できる文字列，発言日時とユーザ ID が重複する発言および，リツイート，機械的に投稿されたと判別できるものを取り除いておく．収集期間は，2015 年 1 月〜2 月とし，統計情報は，表 3 に示すとおりである．感情表現の多くが一般的に用いられ，流行り廃りが少ないと考えたため，短期間の収集とした．本論文では，以降，若者言葉 Tweet コーパスと感情表現 Tweet コーパスの 2 つをあわせて，Tweet コーパスと呼ぶことにする．

5. 若者言葉から標準語への変換手法

本章では，若者言葉を入力とし，その若者言葉に意味的にも感性的にも類似する語を出力する手法について述べる．図 4 に，変換の流れを示す．以下，変換のプロセスを図を用いて順に説明する．

5·1 文脈類似性に基づく類似語の取得

ある若者言葉とその他の語との文脈的な類似性を計算するため，対象となる単語の周辺単語から学習された単語ベクトル（単語の分散表現）を概念ベクトルとして用いる．近年，分散表現を求める幾つかの手法が実装されている [Mikolov 13, Pennington 14].

本研究では，分散表現の学習ツールとして word2vec [word2vec] を用い，単語の skip-gram による学習をおこなう．学習させるコーパスは，前節で説明した Tweet コーパスである．このコーパスを形態素解析にかけ，分かち書きにしたものを使用する．ただし，若者言葉の多くが誤分割されるため，あらかじめ前処理をおこなうことによって，正しい分割がおこなわれるようにした．

入力された若者言葉と，概念ベクトルデータベースに登録されている語との概念ベクトル間の類似度（概念類似度）を，コサイン類似度により計算する．この計算結果から，概念類似度の閾値 T_c 以上の語を類似語集合 $w_j \in SYM$ として取得する．この類似語集合には，関連はするが変換対象には適さない語が多量に含まれてしまうため，後述する俗語らしさの計算によるフィルタリングおよび感性類似度に基づくフィルタリングを適用する．

5·2 俗語らしさの計算

概念ベクトル間の類似度の計算対象は，コーパス内のベクトル化可能なすべての単語である．そのため，変換候補として適さない標準語以外の語も出力に含まれてしまう．標準語のみを類似度計算の対象とする方法も考えられるが，シソーラスや形態素解析用辞書に登録されていない語も存在する．若者言葉を標準語に変換することが目的であるため，俗語のような語を候補から除去したい．そのため，出力された類似語候補のなかから，俗語らしさを評価する指標を用いることで，俗語らしさが高い語を除去または順位を下げる方法をとる．

表 4 文字列から抽出する表層特徴量

No.	特徴量 【種類数】
1	文字種【18 種】
2	文字の画数の総和
3	文字（漢字）の使用用途【2 種】
4	文字（漢字）の修得学年【7 種】
5	文字（漢字）の検定レベル【10 種】
6	文字（漢字）の部首【226 種】
7	文字（漢字）の読み【1296 種】
8	すべての部分文字列の単語親密度【7 種】

従来研究において，若者言葉を，構成文字の表層的特徴に基づき，文中から抽出する手法がある [Matsumoto 14]．しかし，文中からの抽出の場合は，周辺情報も含めた表層的特徴が重要であるが，本研究では，入力は若者言葉そのものである．この場合，入力時の周辺情報は考慮されない．本研究では，俗語の持つ文字特徴に着目し，俗語らしさの数値化をおこなう．また，俗語らしさと同時に，標準語らしさも計算する必要がある．

表 4 に, 文字特徴量として抽出する表層特徴量の種類を示す．No.1〜No.8 までの特徴量をベクトルで表現すると，1,567 次元となり，画数の総和以外は，各特徴の対象文字列における出現頻度を各次元の値とする．俗語，標準語双方の文字列から表層特徴量を抽出し，表層特徴量間の類似度を求めることにより，スコア付けをおこなう．単語親密度には,「日本語の語彙特性」[天野 03, 天野 08] に収録されている単語と，付与されている単語親密度値を用いた．入力される単語 w_i の俗語らしさのスコア $Sc(w_i)$ のスコアを (1) により計算する．$sim(w_i, ys_j)$ は，単語 w_i と俗語 ys_j との表層特徴量間の類似度（表層類似度）を示す．$ys_j = (ys_1, ys_2, \ldots, ys_k)$ は，単語 w_i との表層類似度が上位 k 位までの俗語の集合を示す．同様に，$stw_j = (stw_1, stw_2, \ldots, stw_k)$ は，単語 w_i との表層類似度が上位 k 位までの標準語の集合を示す．標準語らしさのスコアは，$Sc(w_i)$ に -1.0 を掛けたものとなる．$Sc(w_i)$ の値があらかじめ定められた閾値 T_h よりも小さい場合は，w_i を出力候補から除外する．

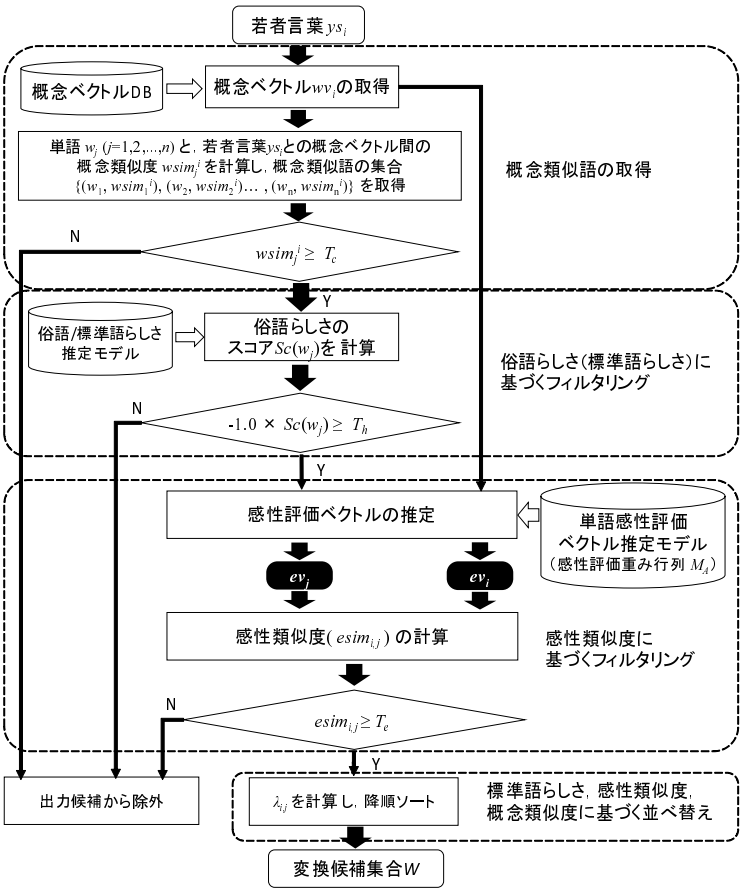


図 4 若者言葉から変換候補集合取得の流れ

$$S(w_i) = \frac{1}{k}\sum_{j=1}^{k} sim(w_i, ys_j)$$

$$H(w_i) = \frac{1}{k}\sum_{j=1}^{k} sim(w_i, stw_j)$$

$$Sc(w_i) = S(w_i) - H(w_i) \quad (1)$$

5·3 感性類似度に基づく候補抽出

若者言葉と類似する標準語候補が多数得られた際，意味的な類似性だけを考慮するのではなく，若者言葉の持つ感性と類似した候補を優先的に提示したい．たとえば，「オサレ」という若者言葉は,「おしゃれ」という標準語と意味的に対応している．しかし,「オサレ」という若者言葉で表現することで皮肉や，卑下，揶揄といったネガティブな意味を含むようになる．本手法では，ポジティブな意味の「おしゃれ」だけではなく，意味的に完全に置換が可能ではないが，ネガティブな意味の変換候補も得られると考える．

単語間の感性的な類似性については，感情ベクトル類似度 [Matsumoto 12b] が提案されているが，この研究で提案されているベクトルは快/不快, 覚醒/睡眠の 2 次元のベクトルである．若者言葉は豊富な感性を表現可能であることは本研究で実施したアンケート結果からも明らかであるため，より複雑な特徴量を用いるべきである．また，熊本ら [熊本 11] の研究では，新聞記事を対象としてテキストの印象を抽出する手法を提案している．この研究では,「楽しい」や「のどか」などの 42 語の印象語をもとに，新聞記事の印象を表現するのに適した 3 本の印象軸（「楽しい⇔悲しい」,「うれしい⇔怒り」,「のどか⇔緊迫」）を決定している．しかし，新聞記事を表現する印象よりも，若者言葉などが持つ単語の感性のほうが，より複雑と考えられる．

本研究では，単語の感性的な類似性を計算するため，感性を表現する特徴量（以下，感性評価ベクトルと記述）を用いる．アンケートから得られた俗語の感性評価値と，その俗語の概念ベクトルとの関連性を計算し，感性評価値が未知の語に対して，概念ベクトルから感性評価ベクトルを求める．

アンケートに用いた評価対 16 種類を感性評価ベクトルとして，アンケート取得済みの若者言葉の概念ベクトルにおける各次元との関連度を求め，その関連度を要素とする感性評価重み行列 $\boldsymbol{M_A}$ を得る．感性評価重み行列 $\boldsymbol{M_A}$ の計算式を，(2) に示す．$\boldsymbol{ev_i} = (e_1^i, e_2^i, \ldots, e_{16}^i)$ は，若者言葉 i の感性評価ベクトルの正規化後の数値を示す．また，$\boldsymbol{wv_i} = (v_1^i, v_2^i, \ldots, v_d^i)$ は，若者言葉 ys_i の d 次元で学習された概念ベクトルの数値を示す．

$$\boldsymbol{M_A} = \sum_{i=1}^{m} \boldsymbol{ev_i^T} \times \boldsymbol{wv_i} = \sum_{i=1}^{m} \begin{pmatrix} e_1^i \\ e_2^i \\ \vdots \\ e_{16}^i \end{pmatrix} \times \left(v_1^i\ v_2^i\ \ldots\ v_d^i\right) \quad (2)$$

感性を未評価の語に対し，d 次元の概念ベクトルを取得し，$\boldsymbol{M_A}$ を掛けることにより，感性評価ベクトル ($\boldsymbol{ev_j}$) の導出をおこなう (3).

$$\boldsymbol{ev_j^T} = \boldsymbol{M_A} \times \boldsymbol{wv_j^T} = \left(e_1^j\ e_2^j\ \ldots\ e_{16}^j\right) \quad (3)$$

感性評価ベクトル推定により，入力された若者言葉 ys_i から推定された感性評価ベクトル $\boldsymbol{ev_i}$ と，変換候補として得られた単語 w_j の感性評価ベクトル $\boldsymbol{ev_j}$ との，感性類似度 $esim_{i,j} = sim(\boldsymbol{ev_i}, \boldsymbol{ev_j})$ を計算する．この値が閾値 T_e よりも小さければ，単語 w_j を変換候補から除外する．また，概念ベクトル類似度 $wsim_{i,j} = sim(\boldsymbol{wv_i}, \boldsymbol{wv_j})$ を cosine 類似度により求め，この値と感性類似度 $esim_{i,j}$ の相加平均値に，俗語らしさのスコア $Sc(\boldsymbol{w_j})$ に -1.0 を掛けた値（標準語らしさのスコア）を掛け合わせて，スコア $\lambda_{i,j}$ を得る (4)．このスコアにより候補の出力順を決定する．

$$\lambda_{i,j} = -1 \times Sc(w_j) \times \left\{ \frac{esim_{i,j} + wsim_{i,j}}{2} \right\} \quad (4)$$

6. 評 価 実 験

6·1 予 備 実 験

提案手法で用いる俗語らしさのスコア計算および，感性評価ベクトル推定手法の評価をおこなう．まず，俗語らしさの計算モデルの評価（予備実験 1）では，訓練データとして，俗語と標準語それぞれ 2,386 語ずつに対し，表層特徴量を登録したデータベースを構築し，俗語らしさのスコア計算をおこなう．

実験において，入力が俗語の場合，俗語らしさのスコアが正の値の場合に正解，負の場合に誤りとした．同様に，入力が標準語の場合，俗語らしさのスコアが負の場合に正解，正の場合に誤りとした．評価用データは，俗語，標準語ともに 671 語を選択した．

また，感性評価ベクトル推定手法の評価（予備実験 2）においては，アンケートに用いた若者言葉に対し，交差検定により，推定された感性評価ベクトルと正解ベクトルとの cosine 類似度に基づき評価する．cosine 類似度が高いほど，モデルの概念ベクトルから感性評価ベクトルへの再現能力が高くなるため，良いモデルであるといえる．

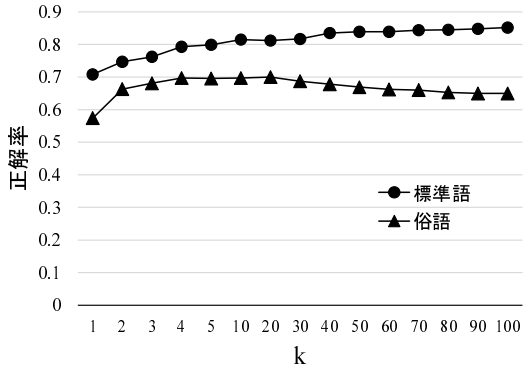


図 5 俗語/標準語判定実験結果（k の値による正解率推移）

また，提案手法では，入力となる若者言葉の変換候補となる標準語に対して，出力順を定めるため，感性評価ベクトル間の類似度（以降，感性類似度と記述）を cosine 類似度計算により求める．そのため，感性評価アンケートにより評価を得ていない標準語に対して，感性評価ベクトル推定モデルによりどの程度妥当な感性評価ベクトルが得られるかの評価が必要である．

標準語のなかには，感情を表現する語も多く存在する．本予備実験では，感情表現に対し，感性評価ベクトルを推定した際に，感性評価ベクトル中の，感情軸と，感情表現の表す感情との一致をみることでの評価もおこなう．

6·2 予 備 実 験 結 果

予備実験の結果を図 5，図 6 に示す．

図 5 は，文字の表層的特徴に基づき，k 近傍法により俗語らしさ/標準語らしさのスコアを計算し，正解率を求めたものである．標準語については，k の値が大きくなるにつれて正解率が上昇する傾向があり，一方で，俗語については，k の値が 20 を境として正解率が低下する傾向があった．本研究では，できるだけ多くの適切な標準語候補を得たいため，標準語が俗語として誤判定されることを避けたい．そのため，本予備実験の結果から k の値は 100 が最適であると判断した．

図 6 は，若者言葉の感性評価ベクトル推定を交差検定により求め，正解ベクトル（アンケートにより得たベクトル）との cosine 類似度を計算し，平均値を得た結果を，概念ベクトルの学習条件ごとに比較したものである．縦軸が cosine 類似度を示している．横軸には，概念ベクトルの学習に用いた window サイズと概念ベクトルの次元数のパラメータの組合せ (window:dimension) を示している．window サイズが 10，概念ベクトルの次元数が 50 のときに最も類似度が高くなった．

また，感情表現辞典において感情カテゴリが決定されている標準語（感情表現）1,071 語の感性評価ベクトルについて推定をおこなった．感情表現辞典において定義されている感情表現が示す感情カテゴリと，推定された

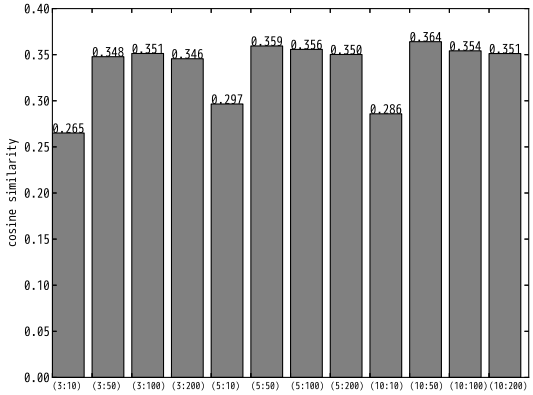


図 6 感性評価ベクトル推定結果と正解ベクトルとの cosine 類似度平均の比較

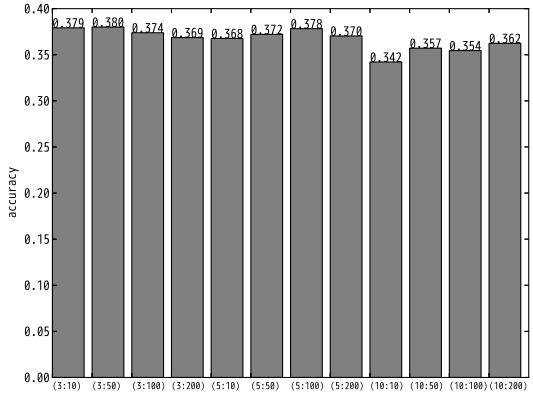


図 7 感情表現に対する感性評価ベクトル推定の正解率

感性評価ベクトルにおける該当する感情カテゴリの感情評価軸の値（正負）が一致している場合を正解，それ以外を誤りとして，正解率を求めた．各パラメータ組合せにおいての正解率を図 7 に示す．

パラメータの組合せによる正解率の差はあまりみられなかったが，window サイズ 3, 次元数 50 の組合せのときに最も高い正解率 0.38 を得た．これらの予備実験の結果から，感性評価ベクトルの推定においては，window サイズよりもベクトルの次元数に影響を受けることがわかる．このため，次節で述べる評価実験では window サイズは 10 で固定し，次元数を (10，50，100，200) の 4 通りについて評価することにした．

6·3 若者言葉の標準語への変換実験

提案手法では Tweet コーパスをもとに学習した概念ベクトルに基づき，入力された若者言葉と概念が類似する単語集合を取得し，そのなかから標準語と判定できるものを選別したうえで，感性的に類似する語に重みづけをして変換候補を出力する．

表 5 若者言葉と対応する標準語の例

若者言葉	標準語
チキン	臆病，弱虫，意気地無，腰ぬけ，小心者
ゲトる	ゲットする，手中に収める，手に入れる，手に入る，得る
ギャル	女の子，女子
キワい	際どい，危うく，すんでのところで，危ない，危なっかしい
コケた	お蔵入り，お流れ，頓挫する
末期的	最悪，最低，どん底
もっさり	やぼったい，田舎臭い，やぼ，むさ苦しい，鈍重
ニート	引きこもり，無職，すねかじり

提案手法による出力結果の妥当性について，以下のような評価をおこなう．実験対象となる若者言葉は，感性評価アンケートに用いた 671 語のうち，正解となる標準語候補を人手によりシソーラス（3·2 節で用いた標準語辞書）から選択できた 190 語とする．この 190 語について，1 つの若者言葉あたり平均して約 3 語の標準語候補が正解候補として付与されている．若者言葉と対応する標準語の例を表 5 に示す．

また，上述の評価対象から外れている若者言葉は，シソーラス上から標準語候補を人手により見つけることができなかったものである．この原因のひとつに，若者言葉の原単語の意味から，大きく変化してしまっている場合がある．また，複数の単語での表現が適している場合には，シソーラスからは正解候補を抽出できなかった．

しかし，本研究で目指すのは，若者言葉を，より一般的な語で言い換えるシステムであるため，言い換え先の語が必ずしもシソーラスに登録されている必要はない．このことから，若者言葉の変換候補として俗語が出力される場合においても評価することを考えた．これらの俗語の類似語候補について，評価用の正解データを取得するため，アンケートを実施した．アンケートは，4 名の被験者に，各俗語に対して，概念ベクトル間の類似度計算により出力された類似語候補をランダムで並べ替えて提示し，被験者が，意味や感性が類似している（関連している）と判断可能な語を複数選択する形式とした．

各俗語の類似語候補の提示数は，類似度上位 20 件とした．アンケート結果から，2 名以上の被験者が選択した俗語の類似語候補を正解候補として決定した．また，被験者の選択頻度ごとの類似語候補数を，表 6 に示す．4 名全員が選択した類似語候補数は 26 種類と少なく，2 名以上が選択した類似語候補数は 1,033 となった．

さらに，人手による標準語候補が決定できなかった語で，今回のアンケートの結果により正解候補が決定でき

表 6 被験者の選択頻度ごとの類似語候補数

選択頻度	類似語候補数
1	1,654
2	748
3	259
4	26

た俗語の数は，274 種類となった．正解とした類似語候補には，標準語以外の語が含まれているため，標準語らしさのスコアによるフィルタリングを適用しない場合の精度評価を行う必要がある．

評価に用いる指標として，(5) に示す MRR(Mean Reciprocal Rank) の平均値を用いる．検索結果のうち正解となる単語が N 個出力された場合，単語 i の出力順位が R_i としたときの R_i の逆数の総和を正解単語数で割った値となる．正解が出力されなかった場合，MRR は 0 の値をとる．MRR が高いということは，正解候補をより上位に出力することに成功していることを意味する．以下，標準語の正解候補を持つ俗語に対して計算した MRR を，MRR(1)，アンケートにより正解候補を決定した俗語に対して計算した MRR を MRR(2) と記述する．

$$MRR = \begin{cases} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{R_i} & (N \neq 0) \\ 0 & (otherwise) \end{cases} \quad (5)$$

また，ベースライン手法として，標準語らしさのスコア計算および感性類似度の計算をおこなわず，Tweet コーパスから学習した概念ベクトルにより，概念ベクトル類似度のみで類似語を変換候補として取得する手法を用いた．実験で用いるパラメータの組合せを，window サイズを 10 に固定し，ベクトルの次元数を (10, 50, 100, 200) の 4 通りとした．標準語らしさのスコアの計算時の近傍数 k の値に 100 を設定し，標準語らしさのスコアの閾値 T_h は 0 に設定し [*1]，感性類似度の閾値 T_e は 0.5 に設定した．また，概念的にほとんど類似しないような語を出力候補から除外するため，概念類似度の閾値 T_c に 0.5 を設定した．提案手法とベースライン手法における閾値設定の違いについて，表 7 にまとめて示す．‘–’ は，閾値を設定しないことを示す．

表 7 提案手法とベースラインにおける閾値設定

	T_c	T_h	T_e
提案手法	0.5	0	0.5
ベースライン	0.5	–	–

6·4 実 験 結 果

実験結果を，図 8 に示す．MRR の平均値は，MRR(1), MRR(2) どちらにおいても，概念ベクトルの次元数を 200 と設定したときに，提案手法による結果がいずれのパラメータの組合せによるベースライン手法を上回った．このことから，提案手法は不要な語を，標準語らしさと感性の類似性の両面からフィルタリングできているといえる．また，提案手法では標準語らしさの高い語ほど優先順位を高くする処理をしているため，MRR(2) が MRR(1) よりも低くなったのは，妥当といえる．

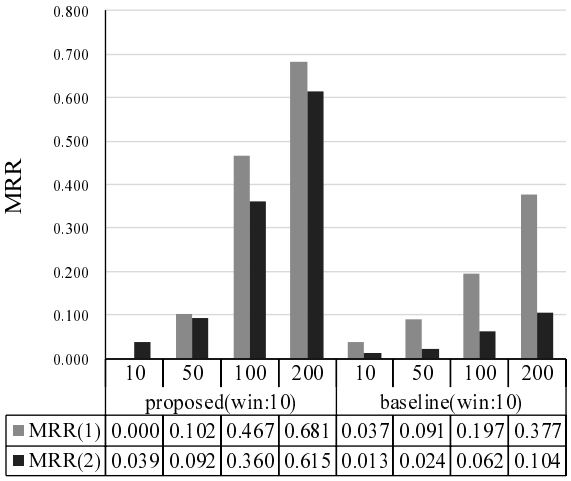


図 8 MRR 平均の比較

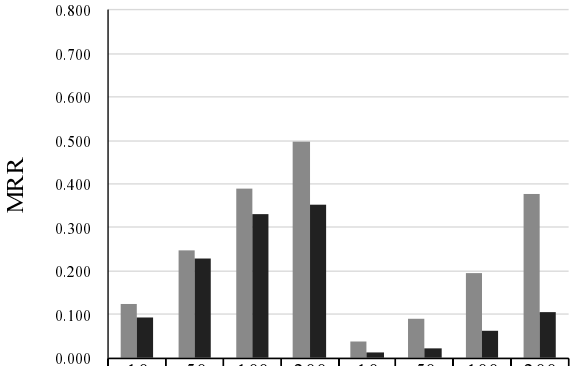
出力された正解候補について，ベースライン手法と提案手法それぞれにおける順位の比較を，表 8 に示す．若者言葉によっては，同義や同じような意味で異表記の単語が多く存在しているため，ベースライン手法による出力順位が下がってしまっていることがわかる．一方で，提案手法は，正解以外の類似語候補をフィルタリングすることで，正解候補の順位を高く保つことができたと考えられる．

ここで，正解の出力数を比較してみたところ，ベースライン手法が最大で 69 語の出力に対して，提案手法は最大で 33 語であり，ベースライン手法を下回るという結果であった．しかし，ベースライン手法では，出力候補数が 100 語以上のものがほとんどであったのに対し，提案手法は，若者言葉 1 語あたり平均約 18 語程度にまで抑えることができていた．

表 8 変換候補の例（window=10, 次元数=100）

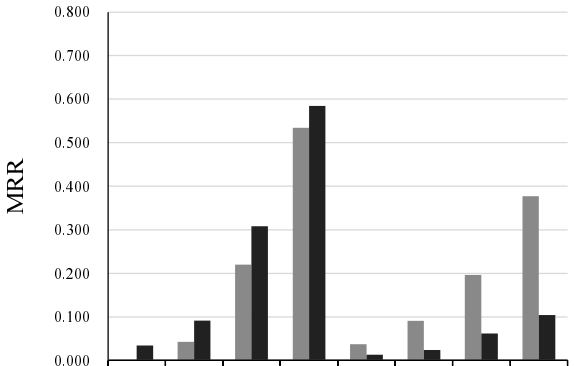
若者言葉	正解候補	正解順位	
		baseline	提案手法
インフル	インフルエンザ	24 位	17 位
いらつく	腹立つ	96 位	3 位
すんごく	ものすごく	16 位	2 位
イケメン	美形	21 位	4 位
インスコ	インストール	77 位	39 位
ていうか	つうか	196 位	1 位
ガン寝	熟睡	79 位	27 位

*1 標準語らしさのスコアが負となる語を除外する．



	proposed(win:10)				baseline(win:10)			
	10	50	100	200	10	50	100	200
MRR(1)	0.124	0.247	0.390	0.498	0.037	0.091	0.197	0.377
MRR(2)	0.094	0.228	0.330	0.354	0.013	0.024	0.062	0.104

図 9 感性類似度のみを適用した場合の比較



	proposed(win:10)				baseline(win:10)			
	10	50	100	200	10	50	100	200
MRR(1)	0.000	0.043	0.220	0.534	0.037	0.091	0.197	0.377
MRR(2)	0.034	0.091	0.308	0.584	0.013	0.024	0.062	0.104

図 10 標準語らしさのスコアによるフィルタリングのみを適用した場合の比較

6·5 考　　察

評価実験の結果，提案手法による類似語のフィルタリングが有効であることがわかった．提案手法がどの程度，感性的に適切な候補を出力可能かについての評価は，今後の検討課題としたい．ここで，標準語らしさのスコアを用いない場合に，出力結果がどのようになるかを調べてみた．図 9 に，感性類似度のみを適用してフィルタリングを施した提案手法について，MRR をベースライン手法と比較したグラフを示す．

次元数が低い（10, 50）ときに，少しではあるが，標準語らしさのスコアによるフィルタリングも適用した場合（図 8）よりも MRR(1) が改善するという結果となった．標準語らしさのフィルタリングを適用した場合には，正解を出力候補に残せないケースがあったと推測される．このことから，「どちらかというと俗語」という判定がされた候補について，概念ベクトルや周辺文脈に基づく判定を加えることで，順位の大幅な低下を回避し，候補にできるだけ残すような改良が必要と考えられる．

つぎに，標準語らしさのスコアによるフィルタリングのみおこない，感性類似度をフィルタリングに適用しない場合の提案手法による結果の MRR をベースライン手法と比較したものを，図 10 に示す．すべての次元数において，MRR(2) のほうが高いという結果が得られている．この結果は，一見，標準語らしさのスコアが高い語が上位に出力されるという予想に反しているように見える．実際に，標準語らしさのスコアによるフィルタリングによって出力候補から除外された語を確認したところ，概念類似語集合のなかには，記号で構成されるような文字列が上位に多数存在している場合があり，そうしたノイズとなる候補をフィルタリングにより多数除去できたことが原因と考えられる．

どちらのフィルタリングも効果を発揮しているといえるが，特に感性類似度については，次元数が少ないときにそれをカバーする効果が大きい．正解候補の出力数を調べてみると，どの次元数の場合においても感性類似度のみを用いた提案手法の方が多くなっていた．このことから，標準語らしさのスコアだけでなく感性類似度を用いることで正解候補の出力漏れを緩和することができると考えられる．

また，若者言葉には様々なタイプの表現があり，表現ごとに適した候補選択の方法が必要と考えられる．たとえば，今回，若者言葉の異表記が多い場合に，MRR が低下するという問題があった．異表記の作りやすさや，作られやすい語については，異表記は異表記として出力するような仕組みを作ることで，標準語ではうまく表現ができない場合に，役立つと考えられる．

7. お　わ　り　に

本論文では，ソーシャルメディア上で多用される若者言葉に着目し，意味と感性の両方の観点に基づく若者言葉の標準語への変換手法を提案した．評価実験の結果，提案手法により，ベースライン手法よりも高い MRR の値を得ることができた．しかし，提案手法では，標準語らしさのスコアおよび感性類似度に基づくフィルタリングにより，正解できる候補数が大幅に減少してしまうという問題がある．また，追加実験により，標準語らしさのスコアを用いず感性類似度のみを適用した場合に，標準語辞書に含まれない正解変換候補の数が増えることも確認できた．

今後は，フィルタリングの精度を上げるため，標準語らしさ/俗語らしさのスコア計算方法を改良したいと考えている．本論文で用いた表層特徴量のみでは，標準語と類似する表記の俗語の影響を受けてしまうため，文脈特徴量も加えて精度改善を検討したい．また，今回用いた訓練データには，標準語と同じ表記を持つ俗語が含まれていた．あらかじめ標準語辞書と照らし合わせて訓練デー

タから除去するなどの前処理も必要である．

提案手法を用いて，発話文中の若者言葉を，感性を保ったまま標準語に変換することができれば，従来の感情推定手法により，若者言葉を含んだ発話文コーパスへの感情タグ付けが容易になると考えられる．今後は，コーパス中の若者言葉の標準語への自動変換を通して提案手法の評価および改良をおこなう予定である．

謝　　辞

本研究の一部は，科学研究費補助金 15K16077, 15K00425, 15K00309 によりおこなわれた．

◇ 参　考　文　献 ◇

[天野 03]　天野 成昭，笠原 要，近藤 公久: NTT データベースシリーズ 日本語の語彙特性 第 1 期 CD-ROM 版 (2003)
[天野 08]　天野 成昭，笠原 要，近藤 公久: NTT データベースシリーズ 日本語の語彙特性 第 4 期 CD-ROM 版 (2008)
[Bond 12]　Bond, F., Baldwin, T., Fothergill, R. and Uchimoto, K.: Japanese SemCor: A Sense-tagged Corpus of Japanese, In Proceedings of the 6th International Conference of the Global WordNet (2012)
[藤田 01]　藤田 篤，乾 健太郎: 語釈文を利用した普通名詞の同概念語への言い換え，言語処理学会第 7 回年次大会発表論文集，pp. 331–334 (2001)
[原田 02]　原田 俊信，亀田 弘之: 若者語の処理方法とその評価，電子情報通信学会技術研究報告. TL, 思考と言語, Vol. 102, No. 491, pp. 1–6 (2002)
[池田 10]　池田 和史，柳原 正，松本 一則，滝嶋 康弘: くだけた表現を高精度に解析するための正規化ルール自動生成手法，情報処理学会論文誌 データベース (TOD), Vol.3, No. 3, pp. 68–77 (2010)
[池原 97]　池原 悟，宮崎 正弘，白井 諭，横尾 昭男，中岩 浩巳，小倉 健太郎，大山 芳史，林 良彦: 日本語語彙大系 CD-ROM 版，岩波書店 (1999)
[情報通信研究機構]　EDR 電子化辞書，情報通信研究機構
[日本語俗語辞書]　http://zokugo-dict.com/.
[鍛冶 13]　鍛冶 伸裕，喜連川 優: 未知語を考慮した形態素解析のための単語ラティスの効率的な生成方法，情報処理学会研究報告. SLP, 音声言語情報処理, 2013-SLP-96, No. 12, pp. 1–8 (2013)
[熊本 11]　熊本 忠彦，河合 由起子，田中 克己: 新聞記事を対象とするテキスト印象マイニング手法の設計と評価，電子情報通信学会論文誌. D, 情報・システム, Vol. 94, No. 3, pp. 540–548 (2011)
[小林 05]　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)
[Kohonen 82]　Kohonen, T.: Self-organized Formation of Topologically Correct Feature Maps, Biological Cybernetics, Vol. 1, pp. 59–69 (1982)
[国立国語研究所 04]　国立国語研究所: 分類語彙表増補改訂版，大日本図書 (2004)
[Matsumoto 11]　Matsumoto, K. and Ren, F.: Construction of Wakamono Kotoba Emotion Dictionary and Its Application, In Proceedings of the 12th International Conference, CICLing2011, Part I, pp. 405–416 (2011)
[Matsumoto 12a]　Matsumoto, K., Kita, K. and Ren, F.: Emotion Estimation from Sentence Using Relation between Japanese Slangs and Emotion Expressions, In Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation, pp. 377–384 (2012)
[Matsumoto 12b]　Matsumoto, K., Kita, K. and Ren, F.: Emotional Vector Distance Based Sentiment Analysis of Wakamono Kotoba, China Communications, Vol. 9, No. 3, pp. 87–98 (2012)
[Matsumoto 14]　Matsumoto, K., Akita, K., Keranmu, X., Yoshida, Y. and Kita, K.: Extraction Japanese Slang from Weblog Data based on Script Type and Stroke Count, Procedia Computer Science, Vol. 35, pp. 464–473 (2014)
[松尾 14]　松尾 朋子，安藤 一秋: 格要素を用いたテンプレートによる若者言葉の自動抽出，情報処理学会第 76 回全国大会講演論文集，pp. 167–168 (2014)
[MeCab]　http://taku910.github.io/mecab/.
[Mikolov 13]　Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, CoRR, abs/1310.4546 (2013)
[森 14]　森 信介，ニュービッグ グラム: 言語資源の追加：辞書かコーパスか，情報処理学会研究報告，自然言語処理研究会報告，2014-NL-216, No. 12, pp.1–3 (2014)
[中村 93]　中村 明: 感情表現辞典，東京堂出版 (1993)
[ニコニコ大百科]　http://dic.nicovideo.jp.
[野口 16]　野口 真人，梶原 智之，小町 守: 語構造情報を用いた日本語複合動詞の言い換え，言語処理学会第 22 回年次大会発表論文集，pp. 729–732 (2016)
[Pennington 14]　Pennington, J., Socher, R. and Manning, C. D.: GloVe: Global Veectors for Word Representation, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP2014), pp. 1532–1543 (2014)
[Ren 15]　Ren, F. and Matsumoto, K.: Semi-automatic Creation of Youth Slang Corpus and Its Application to Affective Computing, IEEE Transactions on Affective Computing, Vol. 7, No. 2, pp. 176–189 (2015)
[三枝 07]　三枝 優一，古井 陽之助，速水 治夫: Web から新語を動的に獲得する形態素解析用辞書拡張方式，情報処理学会研究報告データベースシステム（DBS）, 2007-DBS-141(6), pp. 77–82 (2007)
[佐野 11]　佐野 大樹: 日本語における評価表現の分類体系 : アプレイザル理論をベースに, 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 110, No. 400, pp. 19–24 (2011)
[高村 06]　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, Vol. 47, No. 2, pp. 627–637 (2006)
[土屋 12]　土屋 誠司，鈴木 基之，任 福継，渡部 広一: モーラ系列と音象徴ベクトルによるオノマトペの印象推定法，自然言語処理，Vol. 19, No. 5, pp. 367–379 (2012)
[内田 12]　内田 ゆず，荒木 健治，米山 淳: ブログ記事からのオノマトペ用例文の自動抽出手法，知能と情報，Vol. 24, No. 3, pp. 811–820 (2012)
[Web Corpus]　日本語 Web コーパス，http://s-yata.jp/corpus/nwc2010/.
[word2vec]　word2vec，https://code.google.com/archive/p/word2vec/.
[山西 15]　山西 良典，大泉 順平，西原 陽子，福本 淳一: 人名の言語的特徴の分析に基づくキラキラネーム判定，日本感性工学会論文誌，Vol. 15, No. 1, pp. 31–37 (2015)
[米川 98]　米川 明彦: 若者言葉を科学する，明治書院 (1998)
[若者言葉辞典]　若者言葉辞典：http://bosesound.blog133.fc2.com/.

〔担当委員：奥 健太〕

2016 年 4 月 9 日　受理

著　者　紹　介

松本　和幸

2008 徳島大学大学院工学研究科博士後期課程修了．博士（工学）．2009 年 10 月より現在まで，徳島大学大学院ソシオテクノサイエンス研究部助教．感情計算，自然言語処理，対話処理，知的英作文支援等の研究に従事．情報処理学会，電子情報通信学会，言語処理学会，電気学会，ヒューマンインタフェース学会各会員．

土屋　誠司（正会員）

2000 年同志社大学工学部知識工学科卒業．2002 年同大学

院工学研究科知識工学専攻博士前期課程修了．同年，三洋電機株式会社入社．2007 年同志社大学大学院工学研究科知識工学専攻博士後期課程修了．同年，徳島大学大学院ソシオテクノサイエンス研究部助教．博士（工学）．2009 年同志社大学理工学部インテリジェント情報工学科助教．2011 年同准教授．主に，知識処理，概念処理，意味解釈の研究に従事．言語処理学会，情報処理学会，日本認知科学会，電子情報通信学会各会員．

芋野　美紗子（正会員）

2009 年同志社大学工学部知識工学科卒業．2011 年同大学院工学研究科情報工学専攻博士前期課程修了．2014 年同大学院工学研究科情報工学専攻博士後期課程修了．2016 年 4 月より大同大学講師．主に，概念処理の研究に従事．言語処理学会会員．

吉田　　稔（正会員）

1998 年東京大学理学部情報科学科卒業．2003 年東京大学大学院理学系研究科博士課程修了．博士（理学）．東京大学情報基盤センター助教を経て，2013 年より徳島大学大学院ソシオテクノサイエンス研究部講師．テキストマイニングの研究に従事．情報処理学会，言語処理学会，日本データベース学会各会員．

北　　研二

1981 年，早稲田大学理工学部数学科卒業．1983 年，沖電気工業（株）入社．1989 年，カーネギーメロン大学機械翻訳研究所客員研究員．1992 年，徳島大学工学部講師．1993 年，同助教授．2000 年，同教授．2002 年，同大学高度情報化基盤センター教授．2008 年，同センター長．2010 年より，同大学大学院ソシオテクノサイエンス研究部教授．博士（工学）．言語処理，情報検索，メディア情報学等の研究に従事．1994 年，日本音響学会技術開発賞受賞．著書「確率的言語モデル」（東京大学出版会），「情報検索アルゴリズム」（共立出版）など．情報処理学会，言語処理学会各会員．