

## 宣伝効果を促進するための親しみやすいツイートへの言い換え

佐藤 朋美 †

† 筑波大学 情報学群 知識情報・図書館学類

関 洋平 ‡

‡ 筑波大学 図書館情報メディア系

### 1 はじめに

近年、企業におけるソーシャルメディアの利用率は、増加している [1]。ソーシャルメディアの中でも、Twitter<sup>1</sup> は、コストがかからないため、企業が商品や催物の広告を発信する手段として注目されている [2]。しかし、多くのユーザは広告ツイートには関心を払わないという傾向がある [3]。

本研究では、この課題を解決するために、広告ツイートを親しみやすいツイートに言い換えることにより、広告活動に対する宣伝効果の促進に有効であるかを検証する。Twitter においてユーザが親しみを持つツイートは、(1) 口語的表現、(2) 絵文字や顔文字が多いといった特徴がみられる。本研究では、このような特徴が顕著にみられる若い女性のツイートコーパスを利用して言い換えを行うと共に、若い女性のツイートらしさを推定し、言い換えの優先順位の決定に利用することが、宣伝効果の促進に有効であることを明らかにする。

### 2 関連研究

親しみやすいツイートへの言い換えを行うための、関連研究について紹介する。Shen ら [4] は、口語的な言い回しが増えることで、親しみやすさが向上すると述べている。また、廣瀬ら [5] によると、文章に顔文字や絵文字を付加すると、感情伝達を促進する効果がある。このような、口語的で顔文字や絵文字を利用した文章は、若い女性が積極的に発信する傾向が見られる [6]。

宮崎ら [7] は、言い換え前と言い換え後の文の差分から、言い換え規則を獲得し、言い換え規則集を構築した。本研究でも、被験者による言い換えに基づくコーパスを構築し、言い換え規則を獲得する。

長浜ら [8] は、単語を素性とすることで精度が高いツイートの性別推定を実現できることを明らかにした。本研究では、若い女性のツイートによく現れる単語を素性として若い女性らしさを推定することにより、言い換え規則の優先順位の決定に利用する。

### 3 親しみやすいツイートへの言い換え手法

#### 3.1 親しみやすいツイートの特徴

本研究では、2 節の関連研究をもとに、若い女性にとっての親しみやすいツイートには、以下のような特徴があると考えられる。

- 口語的表現
- 顔文字や絵文字の利用

このような特徴を反映したツイートに言い換えることが、同世代の女性への、宣伝効果の促進に役立つか検証する。そのために、若い女性の特徴を反映したツイートコーパスを作成し、言い換えに利用する。

#### 3.2 言い換え規則集の構築

##### 若い女性らしさを反映したツイートコーパスの構築

本研究では、親しみにくい広告ツイートを、多くの若い女性の被験者に言い換えてもらい、若い女性らしさを反映したツイートコーパスを構築し、言い換え規則の獲得に利用する。

具体的には、Twitter 上の広告で利用されていた 300 ツイートを対象として、18 名の 10 代～20 代女性が 1 人当たり 100 文ずつ言い換えを行う。言い換えられたツイートは 1,800 件となる。言い換えを行う 300 ツイートは、文語的表現であるか、顔文字や絵文字が利用されていないかを、人手で判別し、収集した。被験者には、「自分が企業の公式アカウントを担当し、同じ世代(若い女性)に宣伝する」ことを考慮し、Twitter 上で普段友人とやり取りをするような表現への言い換えを指示する。また、普段 Twitter を利用するデバイスを使用して言い換えを行うよう指示する。

##### 言い換え規則の獲得

上記のコーパスにおける、言い換え前のツイートと、言い換え後の若い女性のツイートの差分から、言い換え規則を獲得し、言い換え規則集を構築する。言い換え前と言い換え後のツイートを MeCab<sup>2</sup> で形態素解析(辞書は mecab-ipadic-neologd/mecab-ipadic-neologd<sup>3</sup> を使用)し、動的計画法を利用して形態素ごとの対応付けを行い、表層表現が一致していない形態素のペア、もしくは形態素の連鎖(*n-gram*)のペアを、言い換え規則として獲得する。結果、獲得できた言い換え規則は、2,019 件、2 回以上獲得された規則は、325 件であった。

#### 3.3 言い換え規則の優先順位の決定

言い換えを行う際に、複数の言い換え規則が適用できる場合があるが、その優先順位を決定する必要がある。本研究では、より若い女性らしい言い換え規則を優先する。そのため、若い女性らしさを推定する推定器を構築する。言い換え規則に対する若い女性らしさの推定値は、言い換え規則を獲得する際の言い換え後のツイートに対する若い女性らしさの推定値とする<sup>4</sup>。

推定器の構築には、LIBSVM<sup>5</sup> を使用した。訓練データは、2015 年 8 月 9 日から 2015 年 9 月 15 日までに 100 件以上ツイートをしたアカウントの中で、プロフィールと過去のツイートから、若い女性のアカウントか若い女性以外のアカウントかを人手で判断できた各 40 アカウントの 100 ツイートずつ、合計 8,000 ツイートである。素性の選択に利用したデータは、訓練データと同条件であり別のアカウント各 40 アカウントの 100 ツイートずつ、合計 8,000 ツイートである。このデータに出現する全単語を利用し、若い女性のツイートについての相互情報量を算出し、上位 3,000 語を素性とした。

Paraphrasing Tweets into Intimate Style to Promote Advertising Effect

†Tomomi SATO ‡Yohei SEKI

†College of Knowledge and Library Science, School of Informations, University of Tsukuba

‡Faculty of Library, Information and Media Science, University of Tsukuba

<sup>1</sup><https://twitter.com/>

<sup>2</sup><http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

<sup>3</sup><https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>

<sup>4</sup>複数のツイートに同じ規則が現れた場合には、規則が現れたツイートの推定値の合計を推定値とする。

<sup>5</sup><https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

10 分割交差検定により評価したところ、正解率は 70.25 % となった。本研究では、この推定器を利用し、言い換え規則の優先順位を決定する。

### 3.4 親しみやすいツイートへの言い換え

親しみやすいツイートへの言い換え方法について説明する。まず、親しみにくいツイートを入力する。次に、入力されたツイートと、言い換え規則との照合を行うために、入力されたツイートの形態素  $n$ -gram ( $n=1 \sim 9$ ) 情報を取得する。より多くの形態素から成り立つ言い換え規則との照合を優先するために、入力されたツイートの長い  $n$ -gram から順に言い換え規則集と照合を行う。最後に、より若い女性らしい言い換え規則を適用した 3 件の言い換え結果を出力する。

## 4 評価実験

### 4.1 方法

提案手法を評価するために、被験者による評価を行う。被験者は 21~22 歳の女性 4 名である。被験者には、親しみにくい広告ツイート 3 件を自由に入力させた。入力すると、1 件につき、若い女性らしさの推定値を優先して言い換えられたツイート上位 3 件が表示される。(1) この言い換え結果 36 件 (入力 3 件 × 出力 3 件 × 4 名) に対して、文法的に違和感 (文法ミス、絵文字や顔文字の不適切な使用など) がないかについて、評価した。(2) また、36 件のうち、文法的に違和感が「全くなかった」「あまりなかった」と回答された言い換え結果について、宣伝効果が促進されたかを評価した。(3) 最後に、言い換え規則の優先順位の決定について、(a) 若い女性らしさの推定値を考慮した場合と、(b) 言い換え規則の頻度のみを利用した場合について、それぞれ上位 3 件の言い換え結果を表示し、どちらの優先順位付けが適切か評価した。

### 4.2 結果

(1) 言い換えられたツイートに、文法的に違和感がないかについての結果を表 1 に示す。(2) また、文法的に違和感が「全くなかった」「あまりなかった」場合に、宣伝効果が促進されたかについての結果を表 2 に示す。(3) 最後に、若い女性らしさの推定値を考慮した場合と、規則の出現頻度を利用した場合は、どちらの優先順位付けが適切か評価した結果を表 3 に示す。

表 1: 文法的に違和感がないか

回答項目	回答数	回答割合 (%)
違和感が全くなかった	14	38.9
違和感はあまりなかった	3	8.3
どちらともいえない	2	5.6
違和感は少しあった	10	27.8
違和感がかなりあった	7	19.4
合計	36	100.0

### 4.3 考察

文法に違和感が「全くなかった」「あまりなかった」場合に、言い換え結果の宣伝効果は「かなり促進された」「少し促進された」という評価が、約 82.4 % となった。この結果から、若い女性のツイートコーパスを利用した親しみやすいツイートへの言い換えは、宣伝効果の促進に有効であると考えられる。また、表 3 の結果から、若い女性らしさの推定器を利用した優先順位付けは、有効であるといえる。宣伝効果が促進されたかと

表 2: 宣伝効果が促進されたか (文法的に違和感を感じられない場合)

回答項目	回答数	回答割合 (%)
かなり促進された	8	47.1
少し促進された	6	35.3
変わらない	3	17.6
あまり促進されなかった	0	0.0
全く促進されなかった	0	0.0
合計	17	100.0

表 3: 言い換え規則の優先順位の決定の適切さ

回答項目	回答数	回答割合 (%)
若い女性らしさの推定値	6	50.0
規則の出現頻度	3	25.0
わからない	3	25.0
合計	12	100.0

いう質問で、「かなり促進された」、「少し促進された」と回答された言い換え結果の例を、以下に示す。

— 宣伝効果が促進された言い換え結果の例 —

開店しました。	→	OPEN(*´▽`*)
ぜひお越しくださいませ。	→	お越しください♥
セール中です。	→	SALE チュウ♥
大変お買い得となっております	→	とってもお買い得です

## 5 おわりに

本研究では、広告ツイートの宣伝効果の促進のために、若い女性のツイートコーパスと、若い女性のツイートらしさを推定する推定器を利用して親しみやすいツイートへの言い換えを行った。その結果、言い換え結果が文法的に違和感を覚えない場合に、約 82.4 % の言い換え結果が、宣伝効果を促進するという結果となった。今後の課題として、文法的誤りの改善が挙げられる。

### 謝辞

本研究の一部は、筑波大学研究基盤支援プログラム (B タイプ)、科学研究費補助金基盤研究 B (課題番号 25280110)、萌芽研究 (課題番号 25540159) の助成を受けて遂行された。

### 参考文献

- [1] 総務省. 平成 26 年通信利用動向調査の結果. <http://www.soumu.go.jp/johotsusintokei/statistics/statistics05a.html> (accessed 2015-12-20).
- [2] 伊藤嘉浩, 高橋優音. 日本企業における SNS を用いたマーケティング戦略: 有効な活用とマネジメント. 山形大学紀要 (社会科学), Vol. 45, No. 1, pp. 91-127, 2014.
- [3] 北村智. ソーシャルメディアにおける情報環境の構成と消費者情報行動の関連に関する研究. 吉田秀雄記念事業財団助成研究集, pp. 43-51, 2013.
- [4] Raymond Shen and Hideaki Kikuchi. Estimation of Speaking Style in Speech Corpora Focusing on Speech Transcriptions. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 2747-2752, Reykjavik, Iceland, May 2014.
- [5] 廣瀬信之, 牛島悠介, 森周司. 携帯電話メールによる感情の伝達に顔文字と絵文字が及ぼす影響. 感情心理学研究, Vol. 22, No. 1, pp. 20-27, 2014.
- [6] 三宅和子. 「規範からの逸脱」志向の系譜—携帯メールの表記をめぐる—. 文学論叢, No. 78, pp. 178-162, 2004.
- [7] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 文節機能部の確率的書き換えによるキャラクター変換. 言語処理学会第 21 回年次大会発表論文集, pp. 277-280, 2014.
- [8] 長浜祐貴, 遠藤聡志, 當間愛晃, 赤嶺有平, 山田考治. ユーザーツイート解析による人物像推定手法の提案と検討. 情報処理学会第 76 回全国大会講演論文集, No. 1, pp. 497-498, 2014.