

図 3: 親しみやすいツイートへの言い換え処理

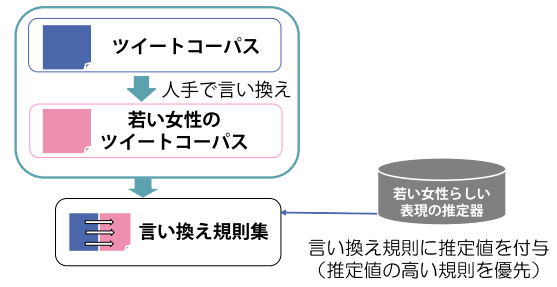


図 4: 言い換え規則集の構築

文章に顔文字や絵文字を付加することで、感情伝達を促進する効果があることを明らかにした。

以上の研究をふまえ、本研究では、若い女性は、口語的な表現で顔文字や絵文字が付加されている文章に親しみをおぼえ、拡散しやすいと仮定する。

2.2 性別推定に関する研究

長浜ら [長浜 14] は、単語を素性とし、高精度なツイートの性別推定を行った。本研究では、若い女性のツイートによく現れる顔文字や絵文字といった単語を素性として、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。

2.3 言い換えに関する研究

宮崎ら [宮崎 14] は、音声対話向け話者キャラクターの言い換えを目的として、文節機能部の言い換え手法を提案している。文節機能部とは、文節から内容語を切り離した部分である。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。

本研究は Twitter の広告拡散力を高める言い換えが目的であるため、文節機能部だけではなく、内容語と文節機能部の言い換えが必要である。本研究では、内容語と文節機能部を言い換え対象として、若い女性である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。

3. 親しみやすいツイートへの言い換え手法

3.1 概要

本研究では、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、口語的な表現や顔文字や絵文字の利用といった特徴があると考えられる。親しみにくいツイートに対して、このような特徴を持つ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ことを検証する。そのため、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。

親しみやすいツイートへの言い換え手法の処理を図 3 に示す。まず、入力された親しみにくいツイートと、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。次に、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、構築した推定器を用いて若い女性らしさの推定値をスコアリングした上位 3 件の言い換え結果を出力する。

3.2 節では、10 代～20 代女性の被験者による広告ツイートの言い換えを通して、若い女性らしさを反映したツイートコーパスを構築し、言い換え規則を獲得する。3.3 節では、若い女

性らしさの推定器を構築し、言い換え後のツイートに対し、言い換え規則の優先順位を決定するための推定値を付与する。

3.2 言い換え規則集の構築

言い換えを行う際に利用する、言い換え規則集の構築方法を図 4 に示す。まず、10 代～20 代の女性の被験者による広告ツイートの言い換えを行い、若い女性のツイートコーパスを構築する。次に、それぞれの言い換え後のツイートに対し、若い女性がよく利用する単語を素性とした推定器を利用し、言い換え規則に若い女性らしさの推定値を付与する。言い換えの際には、推定値の高い規則が優先し出力する。

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、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 ツイートずつ計 1,800 件の言い換えを行う。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、著者が人手で判別し、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した。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した。構築したツイートコーパスの例を、以下に示す。顔文字や絵文字が追加され、より親しみやすいツイートになっていると考える。

いつも A 店をご利用頂きありがとうございます。

いつも A 店をご利用いただき、ありがとうございます (o^o) ♪

いつも A 店をご利用いただきありがとうございます! (*' ω '*)

いつも A 店に来てくれて、ありがとうございます 😊

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。言い換え規則の獲得は、以下の手順で行った。

1. 言い換え前と言い換え後のツイートを MeCab^{*6} で形態素解析

*6 <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

2. 動的計画法を利用した形態素ごとの対応付け

3. 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

4. 以下の項目を出力

- 言い換え前の形態素
- 言い換え後の形態素
- 言い換えられた形態素の品詞
- 言い換えられた形態素の一つ前の形態素の品詞
- 規則として獲得された数 (言い換えられた頻度)
- 若い女性らしさ (推定器の推定値)

手順1では、辞書として mecab-ipadic-neologd^{*7} を使用した。また、言い換え後に現れる顔文字は、1つの単語として扱う必要があるため、顔文字がまとめられているサイト^{*8}のうち、1行で表されている顔文字と、公開されている顔文字辞書^{*9} を利用し、辞書に計 23,100 語の顔文字の追加を行った。

手順2では、言い換え前と言い換え後で、削除・挿入・置換された形態素を対応付けた。また、その際に連続して変更された要素は1つの要素とした。さらに、挿入された要素は、言い換えの際の照合を行うために、言い換え前のツイートの挿入部分の一つ前の形態素も言い換え規則に含むこととした。

なお、「が」を「を」に言い換えるなど、助詞のみの言い換え規則については、文法的に誤りが生じるために除外した。

結果、獲得できた言い換え規則は、2,178 件、1つの形態素に対して2つ以上の言い換え候補が獲得された規則は、317 件であった。獲得結果の例を表1に示す。上の3件は言い換え後の若い女性らしさの推定値が高い要素で、下の3件は低い要素である。

3.3 言い換え規則の優先順位の決定

表1のように、ある形態素に対して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。このような形態素を言い換える場合は、優先順位を決定する必要がある。若い女性にとって親しみやすいツイートへ言い換えるために、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、言い換え後のツイートに対し、若い女性らしさを推定する推定器を構築する。ツイート解析による性別推定は、素性として単語を利用した場合に精度が高いことが報告されている [長浜 14]。本研究では、若い女性と若い女性以外のツイート集合に現れた単語を素性とし、推定器を構築する。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利用する。

推定器の構築には、LIBSVM^{*10} を利用した。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した各40アカウントの100ツイートずつ、計8,000ツイートである。また、素性の選択には、訓練データとは異なる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したアカウント各40アカウントの100ツイート

表 1: 言い換え規則の獲得結果例

言い換え前	言い換え後	若い女性らしさの推定値
です。	! 😊	0.795
しています	中!!	0.503
。	(*^*)	0.475
です。	~	-0.706
しています	!	0.237
。	...	-1.458

ずつ、合計 8,000 ツイートを利用した。このデータに出現する全単語に対し、相互情報量を算出し、若い女性のツイートによく表れるような上位 3,000 語を素性とした。

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。また、構築した推定器を利用して、訓練データを10分割交差検定により評価したところ、正解率は 70.25 % であった。

4. 評価実験

4.1 方法

提案手法の有効性を検証するために、被験者による評価を行った。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。広告ツイートを入力すると、1件につき、若い女性らしさの推定値の上位 3 件の言い換え結果が表示される。

また、被験者 4 名は、言い換え結果 9 件 (入力 3 件 × 出力 3 件) に対して、以下の指標に基づいて評価を行う。なお、指標 1 に関しては、4 名がそれぞれの優先順位付けに基づく出力結果の 3 件に対して、総合的に判断をする。

1. (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、どちらの優先順位付けが適切であるか。
2. 言い換え結果について、宣伝効果は促進されたか。
3. それぞれの言い換え結果について、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) はないか。
4. 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、宣伝効果は促進されたか。

4.2 結果

被験者による評価実験の結果を表 2, 3, 4, 5, に示す。

まず、言い換え規則の優先順位の決定については、50% の結果について、若い女性らしさの推定値を利用した場合の方が、言い換えが適切と評価している。一方、規則の出現頻度を利用した場合は 33.3% であった。このことから、若い女性らしさの推定器を利用した優先順位のスコーリングは、一定の効果があると考えられる。しかし、被験者によっては、特定の絵文字 (例えば、舌が出ているもの) に対して不快感を感じる場合があった。したがって、一部の絵文字を規則から排除することで、より若い女性らしさの推定器のスコーリングの効果が向上すると考える。

また、言い換え後のツイートに対しての宣伝効果の促進に関しては、促進されたという回答が 52.7% であり、促進されなかったという回答が 16.7% であった。言い換え後のツイートに

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

*8 <http://matome.naver.jp/odai/2135442470570037201>

*9 <https://raw.githubusercontent.com/nabokov/mecab-dic-overdrive/master/misc/dic/kaomoji.tsv>

*10 <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 2: 言い換え規則の優先順位の決定について

回答項目	回答数	回答割合 (%)
若い女性らしさの推定値を利用	6	50.0
規則の出現頻度を利用	4	33.3
わからない	2	16.7
合計	12	100.0

表 3: 宣伝効果の促進について

回答項目	回答数	回答割合 (%)
かなり促進された	12	33.3
少し促進された	7	19.4
変わらない	11	30.6
あまり促進されなかった	5	13.9
全く促進されなかった	1	2.8
合計	36	100.0

は、文法的な誤りが生じる場合がある。このようなツイートは宣伝には適さない。そこで、言い換え結果に文法的な違和感を感じるかについて調査したところ、違和感をあまり感じなかった被験者が半数の 50.0% であることに対して違和感を感じた被験者は 30.6% であった。文法的な違和感を解消する点については今後の重要な課題である。

次に、文法的に違和感を感じないと回答があった 18 件の言い換え結果についての宣伝効果の促進に関しては、促進されたという回答が 88.9% であった。これらの結果から、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、文法的に違和感のない場合、宣伝効果の促進に有効であると考える。

最後に、宣伝効果が促進されたかという質問で「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(*'▽`*)
セール中です。	→	SALE チュウ♥
ぜひお越しくださいませ。	→	お越しください
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を目的とし、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とを利用して、親しみやすいツイートへの言い換えを行った。また、被験者実験を通して、言い換え結果が文法的に違和感を覚えない場合には、88.9% の言い換え結果が、宣伝効果を促進するということを検証した。今後の課題としては、文法的誤りを改善することにより、より多くの広告ツイートの宣伝効果を促進することが挙げられる。

謝辞

被験者実験にご協力いただいた皆様のお力により、本論文を完成させることができました。ここに感謝の意を表します。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) の助成を受けて遂行された。

表 4: 文法的な違和感について

回答項目	回答数	回答割合 (%)
違和感が全くなかった	15	41.7
違和感はあまりなかった	3	8.3
どちらともいえない	7	19.4
違和感は少しあった	10	27.8
違和感がかなりあった	1	2.8
合計	36	100.0

表 5: 文法的に違和感を感じない場合の宣伝効果の促進について

回答項目	回答数	回答割合 (%)
かなり促進された	10	55.6
少し促進された	6	33.3
変わらない	0	0.0
あまり促進されなかった	2	11.1
全く促進されなかった	0	0.0
合計	18	100.0

参考文献

- [Fullwood 07] Fullwood, C. and Martino, O. I.: Emoticons and Impression Formation, *Applied Semiotics*, Vol. 19, pp. 4–14 (2007)
- [廣瀬 14] 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)
- [伊藤 14] 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)
- [北村 13] 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)
- [三宅 04] 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って, *文学論藻*, No. 78, pp. 162–178 (2004)
- [宮崎 14] 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性変換, *言語処理学会第 21 回年次大会発表論文集*, pp. 277–280 (2014)
- [長浜 14] 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)
- [Shen 14] Shen, R. and Kikuchi, H.: 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 (2014)
- [総務 15] 総務省: 平成 26 年通信利用動向調査の結果 (2015), <http://www.soumu.go.jp/johotsusintokei/statistics/statistics05a.html> (accessed 2016-01-10)