距離学習に基づくリフレーミング辞書の構築

平野　秀典† 　福本　文代†† 　李　吉屹†† 　郷　健太郎††

† 山梨大学工学部コンピュータ理工学科　〒 400-8510 山梨県甲府市武田 4 丁目 4-37
†† 山梨大学大学院総合研究部工学域　〒 400-8510 山梨県甲府市武田 4 丁目 4-37
E-mail: †{t19cs061,fukumoto,jyli,go}@yamanashi.ac.jp

あらまし　LINE をはじめとするコミュニケーションアプリ利用において，意図しない否定的なメッセージを相手に送信した場合，相手を不快にさせてしまうことがある．このことから，入力メッセージの意味を保持したまま，肯定的な表現へ変換するリフレーミング処理は，円滑なコミュニケーションを図る上で重要となる．本研究では，単語間の入れ替えによる文章のリフレーミングを目的とした変換辞書の構築手法を提案する．変換辞書は，WordNet を用いて構築した．具体的には WortNet の例文を利用し，極性判定と距離学習を用いた類似性判定により作成した．

キーワード　意味解析，極性解析，距離学習，辞書構築

1 はじめに

LINE をはじめとするコミュニケーションアプリの利用において，意図しない否定的なメッセージを相手に送信した場合，相手を不快にさせてしまうことがあり，人間関係に悪影響を及ぼす可能性がある．このことから，入力メッセージの意味を保持したまま，否定的な表現を肯定的な表現へ変換するリフレーミング処理は，円滑なコミュニケーションを図る上で重要となる．否定的な表現を肯定的な表現へ変換する手法にスタイル変換を用いた手法が挙げられる．しかし，スタイル変換の評価指標は原文の属性に依存しない内容を保持するのみで，属性に依存する内容は保持しないことを許容している．スタイル変換が行えているがリフレーミングできていない例として，「Your voice is loud.」を「Your voice is quiet.」と変換する例が挙げられる．この例では極性変換はできているが，原文ではうるさいと言っていた声を変換文では小さいと言ってしまい，文の意味が保持されていない．

本研究では，意味の保持に焦点を当て，単語間の入れ替えによる文章のリフレーミングを目的とした変換辞書の構築手法を提案する．変換辞書を作成する利点に，コミュニケーションアプリ中でユーザが単語を入力した際に即座にリフレーミングの提案を行うことができる点が挙げられる．例えば，ユーザが「Your voice is loud」を入力した場合，「loud」を入力した時に即座に「clear」と修正を行うことを提案できる．また，この際いくつかの修正案を掲示することで，ユーザが最も適していると感じた単語を選ぶことができる．本手法は，二つの単語の意味的類似性判定と，各単語の極性判定の二つの手法からなる．意味的類似性判定を行うモデルは triplet margin loss による距離学習により作成した．極性判定には Twitter のツイートを基としたデータセットである SemEval-2017 Task 4 [7] で学習を行ったモデルを用いた．辞書の構築は，これら二つの手法を組み合わせて行った．本研究の貢献は，以下の通りである．(1) リフレーミングに焦点を当て，単語ベースのリフレーミング辞書を構築した．(2) リフレーミング前後の単語対の判定に，意味的類似性判定と極性判定の二つを組み合わせた手法を提案した．

2 関連研究

2.1 スタイル変換

極性変換の手法の一つにスタイル変換による手法がある．スタイル変換の目的は文章の極性や丁寧さ等の属性を変更しつつ，属性に依存しない内容は変更しないようにすることである．スタイル変換において変換する属性を極性とすることで，極性変換が行える．極性変換の評価指標として，Mir ら [6] や Fu ら [2] の評価指標が用いられる．Fu らの評価指標は，属性に依存しない内容が保存されているかどうか，スタイルの変換が行えているかの二つの評価指標からなり，Mir らの評価指標は，Fu らの評価指標に加えて文章が自然であるかどうかを含めた三つからなる．スタイル変換の代表的な研究として，Gong ら [3]，Li ら [4] の研究がある．Li らは，スタイルが何に属するかのみがラベル付けされたデータセットを用いた教師なし方式による学習手法を提案した．教師なし方式の学習手法である敵対的手法を用いた手法では，優れたスタイル変換文を出力することが難しい．そこで，Li らは文中から変換元のスタイルに関連するフレーズを削除し，変換後のスタイルに関連する新しいフレーズを検索して結合することでスタイル変換を行った．Gong らは，Li らと同様のデータセットに対して強化学習による学習手法を提案した．スタイル変換を行うモデルはアテンション機構に基づくエンコーダ・デコーダモデルにより作成された．

2.2 SentenceBERT

SentenceBERT は，Bidirectional Encoder Representations from Transformers（BERT）[1] を改良したモデルで，文と文の類似度を算出することに優れているモデルである．従来手法の BERT や RoBERTa で文と文の類似度を算出するには，モデルに比較したい文のペアの組み合わせすべてに対して試す必要が

表 1　synset の例

synset 名	difficult.a.01
属する単語	difficult hard
例文	a difficult task
類似 synset	ambitious.s.01
反意語	easy

あり，大幅に時間がかかってしまう．しかし，SentenceBERT では入力文に対してコサイン類似度によって比較できる文の埋め込みを出力するため，全ての文の組み合わせに対して入力を試す必要がないため大幅に時間を短縮できる．出力される文の埋め込みは，単純に BERT の出力層や CLS トークンの出力を利用した文の埋め込みよりも優れており，文に対して埋め込みを作成する点においても優れている．

2.3　RoBERTa

RoBERTa は，BERT の事前学習の手法に改善を行ったモデルである．BERT では文章にマスク処理を行い学習を行うが，RoBERTa では一つの文章に対してマスク位置を変えた文を複数作成することで，学習データの数を増加させている．また，BERT では入力された二つの文が連続する文であるかどうかを予測する Next Sentence Prediction（NSP）タスクにより学習を行っている．しかし，実験の結果，本タスクの有効性が見られないことが判明したため RoBERTa の学習では NSP タスクを用いなかった．また，バッジサイズを非常に大きくして学習を行う．これらで学習を行うことで，従来の BERT と比較してより優れたモデルになる実験結果が得られた．

2.4　ALBERT

ALBERT は BERT を軽量化したモデルである．BERT ではモデルが非常に大きくなり，必要なメモリ量と学習時間の増大が課題である．そこで，学習時間を短縮した．ALBERT では，単語の埋め込みの次元数を削減し，入力の際に線形変換により次元数を増やすことで軽量化を行った．また，トランスフォーマー層のパラメータを共有して扱うことで，一つのトランスフォーマー層を疑似的な多層のトランスフォーマー層として扱うことで必要なメモリ量を削減した．実験の結果，BERT に対して削減したパラメータ数でより優れたモデルが得られた．

3　WordNet

WordNet [5] は英語の意味辞書であり，各単語が synset と呼ばれる同義語のグループに分類されている．本研究では，意味的類似性判定と極性判定に synset に属する単語，例文，類似 synset，反意語の情報を用いるため WordNet を利用した．例文は，synset に属する単語を用いた例文であり，類似 synset は synset に意味が類似している synset の集合，反意語は反対の意味を持つ単語である．例文，類似 synset，反意語は synset によっては存在しない場合と複数存在する場合の両方がある．

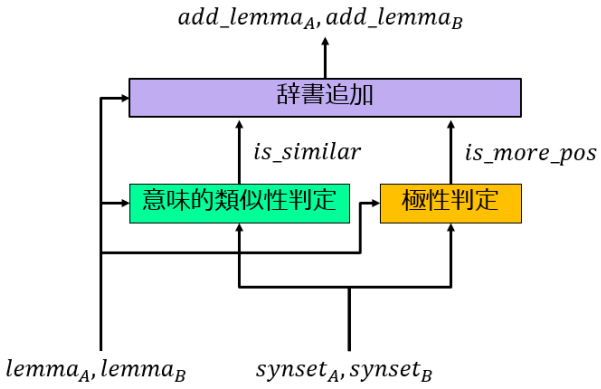

図 1　本手法の概要

表 2　作成文の例

$synset_A$ の例文	a difficult task
$sentence_A$	a hard task
$sentence_B$	a ambitious task

4　提案手法

図 1 は本手法の全体図である．本手法は，図のように三つの要素で構成される．入力は二つの synset の $synset_A, synset_B$ と二つの単語の $lemma_A, lemma_B$ である．$synset_A, synset_B$ の組み合わせは WordNet に存在する形容詞 synset のペアで，$lemma_A$ は $synset_A$ に属する任意の単語，$lemma_B$ は $synset_B$ に属する任意の単語である．出力の add_lemma_A, add_lemma_B は実際に辞書に追加する単語のペアである．

4.1　意味的類似性判定

$lemma_A, lemma_B$ の意味的類似性判定を行う．入力は二つの synset の $synset_A, synset_B$ と二つの単語の $lemma_A, lemma_B$ で，出力は $lemma_A, lemma_B$ の意味的類似性を示す $is\_similar$ である．意味的類似性判定では，$lemma_A, lemma_B$ が用いられた二つの文を作成し，作成文をモデルへ入力する．作成文は，一つの単語列のみが異なり他の部分は全て同じ文章で，異なる単語としてそれぞれ $lemma_A, lemma_B$ が用いられている文章である．この二つの作成文が類似しているならば $lemma_A, lemma_B$ は同じ意味であるとし，類似していないならば $lemma_A, lemma_B$ は異なる意味であるとする．作成文の作成には，$synset_A, synset_B$ の例文を用いる．まず，$synset_A$ の例文の中から，$synset_A$ に属する単語をそれぞれ $lemma_A$ と $lemma_B$ に置き換えて二つの作成文 $sentence_A$ と $sentence_B$ が作られる．実際の作成文の例が表 2 である．複数の例文が存在する場合は，全ての例文に対して同様の処理を行う．

類似性判定を行うモデルには事前学習済みモデルの SentenceBERT をファインチューニングしたものを用いた．SentenceBERT の入力は二つの作成文で，出力は入力の二つの作成文の意味が似ているならば高いコサイン類似度が出力され，二つの作成文の意味が似ていないならば低いコサイン類似度を出力するモデルである．SentenceBERT により得られたコサ

表 3　三つ組の例

a	a difficult task
p	a ambitious task
n	a easy task

イン類似度が大きければ $is\_similar$ は YES となり，低ければ NO となる．なお，$lemma_A, lemma_B$ に対して複数の作成文の組が存在するため，多数決により $is\_similar$ を判定した．SentenceBERT のファインチューニングは，損失関数に式 (1) の triplet loss を用いた距離学習により行った．

$$L(a,p,n) = max\{d(a_i,p_i) - d(a_i,n_i) + margin, 0\} \quad (1)$$

ここで，(a,p,n) を三つ組と呼び，それぞれアンカー入力，ポジティブ入力，ネガティブ入力である．a は基準となる入力で，任意の synset に属する任意の単語を一つ選び，その単語が用いられている文である．p は a で定めた synset の類似 synset を抽出し，その synset に属する任意の単語が用いられている文である．n は a で定めた synset の反意語が用いられた文である．実際に作成した三つ組の例が表 3 である．$d(x,y)$ は x,y の距離を表し距離関数にはコサイン類似度を用いた．$margin$ はマージンを表すハイパーパラメータである．学習では，特徴空間上の a,p 間の距離と a,n 間の距離の差がマージン以上の距離になるようにモデルが更新される．

4.2　極 性 判 定

極性判定の入力は二つの単語 $lemma_A, lemma_B$ で，出力は $lemma_B$ が $lemma_A$ に対して，より肯定的な単語になっているかを表す is_more_pos である．文章の極性判定を行うモデルには，RoBERTa にファインチューニングを行った事前学習済みモデルを用いた．RoBERTa は，入力文に対して極性を肯定的，中立的，否定的の三値で出力する分類モデルである．RoBERTa は，Rosenthal ら [7] の作成した Twitter のツイートに対して肯定的，中立的，否定的のいずれかのラベルを付与したデータセットを用いてファインチューニングを行った．is_more_pos は，$lemma_A, lemma_B$ の極性を算出し，極性が肯定的になっていれば YES，なっていなければ NO とした．極性は，$lemma_A, lemma_B$ が用いられている作成文に対して RoBERTa を用いて極性判定を行い，例文が肯定的なら+1，否定的なら-1 を行った総和である．

4.3　辞 書 追 加

最後に，意味的類似性判定と極性判定の出力を利用して $lemma_A, lemma_B$ のペアを変換辞書に追加するかを判定する．リフレーミングは同じ意味を持ち極性が肯定的に変化していることが条件なため，$(is\_similar, is\_more\_pos) = (YES, YES)$ の場合に $lemma_A, lemma_B$ のペアを辞書に追加する．

5　実　　験

5.1　データ，及び評価尺度

意味的類似性判定と極性判定の評価を行う．意味的類似性判定では，距離学習の際に作成した三つ組のテストデータを評価に用いる．作成された三つ組のうち 80 %は学習データとして学習に用い，20 %を実験に用いるテストデータとした．まずは三つ組を分解し，意味的類似しているペアとしてアンカー入力とポジティブ入力のペアを用いる．意味的類似していないペアとしてアンカー入力とネガティブ入力のペアを用いる．それぞれ 4,000 個のデータに対してコサイン類似度が閾値以上なら意味的類似しているとし，閾値未満なら意味的類似していないとして意味的類似性を判定した．$Acc_{Similar}$ は意味的類似しているペアに対する正答率，$Acc_{NotSimilar}$ は意味的類似していないペアに対する正答率である．Acc は全てのペアに対する正答率である．極性判定では，ファインチューニングに用いたデータセットのテストデータを評価に用いる．Acc は，肯定的，中立的，否定的のラベルが付与されたデータに対して正しく極性のラベルを予測できた正答率である．比較するモデルは，ALBERT と，単純な三層 Neural Network（三層 NN）である．三層 NN の入力はベクトルである必要があるため，SentenceBERT を用いて文をベクトル化したものを入力とした..

表 4　意味的類似性判定の実験結果

閾値	$Acc_{Similar}$	$Acc_{NotSimilar}$	Acc
0.1	0.905	0.563	0.734
0.2	0.862	0.668	0.765
0.3	0.811	0.716	0.763
0.4	0.739	0.801	0.770
0.5	0.629	0.834	0.732
0.6	0.511	0.881	0.696
0.7	0.361	0.932	0.647
0.8	0.211	0.961	0.586

5.2　実　　験

表 4 は意味的類似性判定の実験結果である．最も優れている閾値でも Acc は 77 %で二値分類としては正答率が低くなった．これは，本手法で入力としている作成文は多くが 5 単語程度からなる短文であることが理由だと考えられる．これは，短文に対する問題は長い入力が与えられる問題に対して予測に使える文章情報が少ないためである．また，作成文は例文を基に単語の入れ替えにより作成しているが，作成文は意味の通らない文章である場合もあり，学習データセットとしての質が高いと言えない点も原因だと考えられる．

極性判定の実験結果が表 5 である．三値分類では一般的により高い正答率となることが多いが，Twitter のツイートは予測の難しいデータセットであるため低い正答率になったと考えられる．これは，ツイートは短文が多く，砕けた話し方やハッシュタグ等の SNS 特有の表現が存在するため予測が難しいためである．

5.3　構築した辞書

表 6，表 7 はそれぞれ acceptable，difficult，rich に対して構築された辞書の一部である．実際には，acceptable に対して 30 個，difficult に対して 60 個，rich に対して 3 個の単語が追

表 5　極性判定の実験結果

モデル	*Acc*
RoBERTa	0.715
ALBERT	0.701
三層 NN	0.668

表 6　辞書に追加された単語対（正しく追加されている例）

（acceptable，right）	（acceptable，favorable）
（acceptable，good）	（acceptable，competent）
（acceptable，quality）	（difficult，tight）

加されているが，本表では一部を抜粋している．表 8 は，構築された辞書に対する人手による評価である．acceptable は許容できるという意味の単語で，最低限は満たしているという否定的な捉え方ができるためリフレーミングの余地がある単語だと考えられる．表 6 では，right や favorable 等のリフレーミングとして適した単語が追加されており，正しく単語が追加されている．一方，difficult，rich には意味の類似した単語があまり見られなかった．また，acceptable,difficult では辞書に追加されている単語数が多い．しかし，辞書に追加されている単語数が多いと，リフレーミング候補として多数の単語が候補に挙げられてしまい，ユーザが混乱してしまう．そのため，一つの単語に対して構築される辞書の単語数を減らす必要がある．辞書に追加する単語数を減らすには，意味的類似性判定の閾値を大きくすればよい．しかし，表 4 から分かるように，閾値を上げるほど実際に意味が類似している単語対に対しての正答率が下がってしまい，正しく辞書を構築することができなくなってしまうため，単純に閾値のみをあげることは不適切であると考えられる．辞書に追加されている単語数が多い単語が多い一方で，rich では辞書に追加されている単語数が少ない．これは rich が肯定的な単語であるためリフレーミングする必要がない単語であるためである．

表 8 は，辞書に追加された単語対のうち，50 個の単語対を抽出して筆者が正しく辞書が構築されているかを評価したものである．結果は僅か 12 %と非常に低い値となった．これは，全ての単語に対してリフレーミング前後の単語対かどうかの判定を行っていることが原因である．本辞書で扱っている単語の数は約 3 万個のため，一つの単語に対して 3 万回のリフレーミング判定を実施している．すると，意味的類似性判定や極性判定の正答率が非常に高精度でも，僅かな誤りで多くの単語対がリフレーミング前後の単語対であると判定されてしまうため，辞書に不適切な単語対が多く追加されてしまった．また，リフレーミング候補が存在しないことが妥当である単語に対して多くのリフレーミング候補を辞書に追加してしまっていた．リフレーミング候補が存在しないことが妥当な単語として，既に肯定的な単語や，単純に物事の修飾を行っており極性の観点で見る必要が小さい単語が挙げられる．これらの問題の改善は，意味が類似している可能性の高い単語対に絞って判定を行う手法の導入により行えると考えられる．

表 7　辞書に追加された単語対（誤って追加されている例）

（acceptable，royal）	（acceptable，comfortable）
（difficult，incapable）	（difficult，broken）
（difficult，erratic）	（rich，buttery）

表 8　構築した辞書の評価

リフレーミング候補として適切な単語対	12 %
リフレーミング候補として不適切な単語対	88 %

6 ま と め

本研究ではリフレーミング処理を目的とした変換辞書の構築手法を提案した．リフレーミング前後の単語であるかどうかを，二つの単語間の意味が類似していて，より肯定的な単語になっているかどうかとして判定した．意味が類似しているかの判定は Sentence BERT を用いて，極性判定には RoBERTa を用いた．実験の結果，意味が類似していない単語対も抽出された．今後は，絞り込みの処理についてさらに検討する必要がある．さらに、構築した辞書を用いて文のリフレーミングを行うことも今後の課題として挙げられる．

謝　　辞

本研究は，科研費 20K11904 の支援を受けて実施したものです．

文　　献

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Z. Fu, X. Tan, N. Peng, D. Zhao, and R. Yan. Style transfer in text: Exploration and evaluation, 2017.

[3] H. Gong, S. Bhat, L. Wu, J. Xiong, and W.-m. Hwu. Reinforcement learning based text style transfer without parallel training corpus. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3168–3180, June 2019.

[4] J. Li, R. Jia, H. He, and P. Liang. Delete, retrieve, generate: a simple approach to sentiment and style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1865–1874, June 2018.

[5] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[6] R. Mir, B. Felbo, N. Obradovich, and I. Rahwan. Evaluating style transfer for text. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 495–504, June 2019.

[7] S. Rosenthal, N. Farra, and P. Nakov. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Aug. 2017.