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24回)
SIG-AM-24-07

深層学習を用いた好意と悪意を含む表現の可視化による文章推敲支援

Document Polishing Support by Visualizing Favorable and Malicious Expressions using Deep Learning

庵 翔太[1*]　砂山 渡[2]　畑中 裕司[2]　小郷原 一智[2]
Shota IHORI[1]　Wataru SUNAYAMA[2]　Yuji HATANAKA[2]　Kazunori OGOHARA

[1] 滋賀県立大学大学院工学研究科
[1] Graduate School of Engineering, The University of Shiga Prefecture
[2] 滋賀県立大学工学部
[2] School of Engineering, The University of Shiga Prefecture

Abstract: In recent years, the number of people who interact online has increased due to the increase of SNS users. Considering this situation, it can be said that it is important to revise sentences in consideration of the reader's emotions during dialogue. Therefore, the purpose of this research is to support the polish of sentences that give readers a favorable impression. For this purpose, we implemented functions to visualize expressions including good and bad intentions using deep learning and functions to encourage sentence improvement. As a result of the experiment, it was found that this system can support polish of sentences that give readers a favorable impression.

1　はじめに

近年，スマートフォンの急速な普及に伴うソーシャルメディアの利用者の増加によって，増々多くの人々が，オンラインでのコミュニケーションを行うようになっている．しかしながら，こういったオンラインでの対話では，書き手の表情や細かいニュアンスが伝わらないため誤解が生じやすく，相手の気分を害していまい，トラブルの原因となることがある．また，昨今ソーシャルメディアにおける不適切投稿によって生じるトラブル，いわゆる「炎上」が問題となっている [1].

こういった現状を考えると，オンラインでのコミュニケーションを行う際に，読み手の感情を考慮した文章作成を行うことは，良好な人間関係を構築する上で重要であると言える．例えば，メールを作成する時に，そのメールを読む人の気分を害するような表現を避けることは我々が日常的に行っていることである．そこで，読み手が好感を抱く文章の作成を支援する事が出来れば，コミュニケーションの支援として有用となる．例えば，Twitter などのソーシャルメディア上で，読み手の気分を害する投稿をしようとする書き手のユーザに注意喚起するサービスや，メールの自動添削システムが考えられる．

そこで本研究では，メールやメッセージアプリ，ソーシャルメディアなどのオンラインでの他人とのコミュニケーションにおいて，他人に送りたい文章を入力として，読み手に好感を持ってもらえる文章推敲を支援するシステムを構築する．筆者の学部時代の研究 [2] では，書き手の好意や悪意を表す単語を可視化する機能，文章修正を促すメッセージを表示する機能，文章をスコア付けして表示する機能を実装し，文章修正を支援する．本研究では，これらの機能に加えて，深層学習を用いて，文章の書き手の好意や悪意を含む文を抽出し可視化する．これによって書き手に文章推敲を促し，読み手から好感を持たれる文章の作成を支援することを目的する．

以下本論文では，2 章で関連研究について述べる．3 章で提案する文章推敲支援システムについて述べる．4 章で提案システムの効果を検証した評価実験について述べ，5 章で本論文を締めくくる．

2　関連研究

文章作成を支援する研究は盛んに行われている．

段落の一貫性を数値的に評価する研究がある [3]．この研究では，段落一貫度を定義し，論文執筆などの際

*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻 庵翔太
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: of23sihori@ec.usp.ac.jp

の文章校正支援を期待している．また，文の接続関係を利用する論理方向を可視化し，段落中の話の展開を確認しながら執筆可能な支援環境を用意する研究がある [4]．さらに，英語文章作成時に，過去の添削履歴を提示することにより，学習者に変更を促す英語文章作成支援システムを構築する研究がある [5]．

また，文章から読み手の感情を予測する研究も行われている．

システムがユーザの感情を推測し，その時々の感情状態に応じて感情表現を用いながら，ユーザの説得を行う対話システムを提案する研究 [6] が行われている．この研究では，感情表現が対話に与える影響に着目して，人を説得するための効率的な方法として感情表現を用いている．

また，オンライン上の対話において聞き手の感情を予測し喚起させる研究 [7] がある．この研究では，マイクロブログから大規模な感情タグ付き対話コーパスを構築し，感情喚起モデルを学習することで読み手に感情を喚起させる応答を生成している．

本研究では，他人とのコミュニケーションを目的とした文章を，読み手に好感を持ってもらえる文章に推敲する支援を行うために，深層学習を用いて，文章中の書き手の好意や悪意を表す表現を抽出し可視化するインタフェースを提案する．

3 好意と悪意を含む表現の可視化による文章推敲支援システム

3.1 システムの概要

本研究における文章推敲支援システムの構成を図 1 に示す．本研究では，まず文章を入力として，入力に含まれる好意や悪意を表す単語と，好意や悪意の表現を含む文を抽出する．そして，抽出した単語や文をそれぞれ文章推敲支援インタフェースで可視化する．そのうえで，文の好意悪意度合い表示，文章推敲を促すメッセージ表示，文章のスコア表示，文章の編集機能を提供し，ユーザに文章の推敲を促す．

3.2 好意と悪意を表す単語の抽出 [2]

文章中の好意を表す単語，悪意を表す単語を抽出するために，好意，悪意を表す単語の単語辞書を用いる．単語辞書を構築するために，Twitter 上のツイートとリプライのデータ集合から「喜び」または「悲しみ」を表すリプライの元ツイートに含まれる高頻度語を抽出する．詳細は筆者の学部時代の研究 [2] を参照されたい．

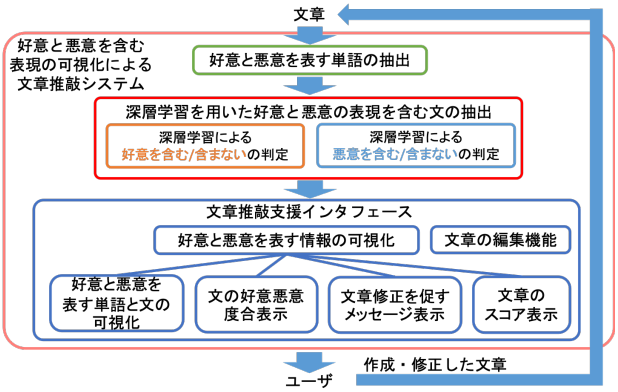

図 1: 文章推敲支援システムの構成

3.3 深層学習を用いた好意と悪意の表現を含む文の抽出

3.3.1 文の抽出に用いる深層学習モデル

文章中の好意の表現を含む文，悪意の表現を含む文を抽出するために，深層学習を用いて文中の単語の組み合わせや出現順などを考慮した文の判定を行う．具体的には，文章中の各文が，好意の表現を含む文か，含まない文かを判定する深層学習のモデル（以下これを「好意モデル」と呼ぶ）と悪意の表現を含む文か，含まない文かを判定する深層学習のモデル（以下これを「悪意モデル」と呼ぶ）の 2 つのモデルを学習し，文の判定を行う．

2 つのモデルの学習には，深層学習の手法の一つである LSTM(Long short-term memory) を用いる．LSTM とは，時系列データの分類や回帰をするモデルである RNN(Recurrent Nueral Network) を記憶を長期にわたって保つように拡張された，長短期記憶と呼ばれる時系列データに対する再帰的モデルの一種である．これを用いることで，単語の組み合わせや出現順を考慮して，各文が好意や悪意の表現を含む文か，含まない文なのか評価する．

3.3.2 学習データの収集

深層学習の学習データには，Twitter におけるユーザ間の対話情報に筆者が人手でラベル付けしたものを 200 次元の分散表現にしたものを使用する．まず，Twitter Streaming API，Twitter REST API[8] を利用することにより，Twitter 上でのある発話 (これを「ツイート」という) について，その発話に返答 (これを「リプライ」という) が行われているツイートを収集する．収集の際には，Twitter 特有の表現を含むツイートは削除する．具体的には，「【定期】」のように使われる「【】」を含む

ツイートは削除する.「【定期】」のような文字列は，定期的に同じ内容のツイートを発信する際に用いられることから，何らかの宣伝や告知のために用いられる事が多い．こういったツイートには書き手の好意や悪意は含まれないことから今回収集するツイートからは削除する．また，ツイート中の URL や記号，絵文字は削除する．これは，URL は文章の上で意味をなさない文字列であるため，書き手の好意と悪意を表す単語にはなり得ないからである．また絵文字や記号は単語ではないため，今回収集したツイートからは削除する．さらに，Twitter にはボット (特定のキーワードに反応し自動返信するプログラム) と呼ばれるユーザが存在する．ボットの発話を収集すると発話と応答の内容に偏りが生じるため，ツイートのユーザ名に「bot」が含まれるツイートまたはリプライは削除する．

3.3.3 学習方法

好意モデルと悪意モデルを学習するための学習データには，3.3.2 項で取得した Twitter 上の対話データのリプライに，筆者が人手でラベルをつけたものを使用する．好意モデルの正例/負例のラベル付けの基準として,「このメッセージをもらった時に嬉しい気持ちになるか/ならないか」を判断する．前者のラベルをつけたデータは，好意の表現を含む文，つまり好意モデルの正例とする．後者のラベルをつけたデータは，好意の表現を含まない文，つまり好意モデルの負例とする．悪意モデルのラベル付けも同様に「このメッセージをもらった時に不快な気持ちになるか/ならないか」を判断する．前者のラベルをつけたデータは，悪意の表現を含む文として悪意モデルの正例とする．後者のラベルをつけたデータは，悪意の表現を含まない文として悪意モデルの負例とする．好意モデルの正例/負例の数はそれぞれ 681 データ，悪意モデルの正例/負例の数はそれぞれ 519 データを用いる．

こうして収集した正例と負例を単語に分かち書きし，各単語を 200 次元の分散表現にする．単語の分かち書きには，オープンソース形態素解析エンジンである MeCab を利用する．また，単語を分散表現にするために，単語分散表現モデルである hottoSNS-w2v[9] を利用する．hottoSNS-w2v は SNS や Web 上の文書を学習コーパスとして word2vec により学習させた単語分散表現モデルである．これにより，学習データの各単語を 200 次元の分散表現とする．最終的な各学習データの形状は [全データ中の最大単語数以上の数, 単語の次元数] となるように [150, 200] のベクトルとする．各学習データの単語数が 150 に満たない場合は，満たない分を 0 で埋める．これを学習データとして，好意モデル，悪意モデルの学習を行う．学習には，好意モデル，悪意モデル共に表 1 のパラメータを使用し，学習データの内 2 割をテストデータとして用いる．

表 1: 好意モデル，悪意モデルの学習パラメータ

エポック	100
バッチサイズ	256
中間層	1 層
中間層ノード数	50
活性化関数	relu
最適化アルゴリズム	Adam
学習率	0.001
ElasticNet 正則化	0.001
ドロップアウト	0.5

表 2: 好意モデル学習結果

	Precision	Recall	F1
正例	0.85	0.75	0.80
負例	0.79	0.88	0.83
平均	0.82	0.82	0.82

表 3: 悪意モデル学習結果

	Precision	Recall	F1
正例	0.74	0.81	0.78
負例	0.85	0.80	0.82
平均	0.81	0.80	0.80

好意モデル，悪意モデルの判定精度を表 2，表 3 に示す．本システムでは,「好意を含まない文」を「好意を含む文」と判定することを避けるために，好意モデルは正例の Precision が高いモデルを採用した．また同様に,「悪意を含む文」を「悪意を含まない文」と判定することを避けるために，悪意モデルは正例の Recall が高いモデルを採用した．上記の手法で学習した深層学習モデルを好意モデル，悪意モデルとし，入力文章の各文が好意，悪意の表現を含む文か，含まない文かをそれぞれ判定する．

3.4 文章推敲支援インタフェース

抽出した好意，悪意を表す単語や文を可視化し，ユーザに文章推敲を促すためのインタフェースを実装する．また，本システムはテキストデータのための総合環境 TETDM(Total Environment for Text Data Mining)[10][11] 上で構築する．図 2 はシステムの表示例であり，システムに文章を入力した時の出力を表示している．また，便宜上各機能を表示している部分を色付きの枠で囲い番号を割り振っている．

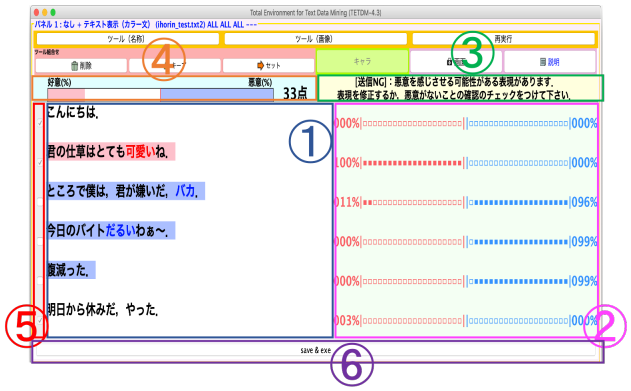

図 2: 文章推敲支援インタフェースの構成

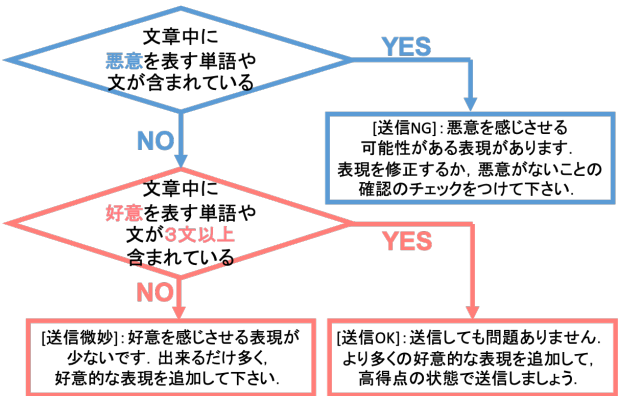

図 3: メッセージ表示までの流れ

3.4.1 好意と悪意を表す単語と文の可視化

表示例を図 2 の①に示す．3.2 節の手法により抽出された書き手の好意を表す単語は文字色を赤色に，悪意を表す単語は文字色で青色で表示している．また，3.3 節の手法により抽出された書き手の好意を含む文は薄い赤色で，悪意を含む文は薄い青色でハイライトしている．文のハイライトについては，3.3.1 項で前述した好意モデルによって，入力文章の各文が好意を含むと判定されると，該当の文を薄い赤色でハイライトし，悪意モデルによって悪意を含むと判定されると，薄い青色でハイライトする．好意にも悪意にも判定された場合は，悪意の判定を優先して薄い青色でハイライトする．この単語の文字色の変更と文のハイライトによって，書き手の好意や悪意を表す単語や文をユーザに視覚的に提示する事で，文章推敲を促す．

3.4.2 文の好意悪意度合い表示

表示例を図 2 の②に示す．文章の各文がどの程度読み手に好意を与えるか，また悪意を与えるか，その度合いを表示する。3.3 節で構築した構築済みの深層学習モデルである好意モデル，悪意モデルによって各文が好意を含む文として判定される確率，悪意を含む文として分類される確率を算出し，それぞれ薄い赤色と薄い青色の棒グラフで可視化する．好意と悪意を表す棒グラフの原点はそれぞれ左右の端であり，100 %の点は棒グラフの中央である．例えば，図 2 において②を見ると，３文目，４文目，５文目の悪意度合いがそれぞれ 96 %，99 %，99 %となっている．これは，３文目，４文目，５文目が読み手に悪印象を持たれる可能性が高い文となることを表している．このように各文が読み手にどんな印象を与えるかを視覚的に表示することによって，文章推敲を促す．

3.4.3 文章推敲を促すメッセージ表示

表示例を図 2 の③に示す．この機能では文章の内容に応じて図 3 の流れに従って，ユーザに文章推敲の指針をメッセージとして表示する．これらのメッセージによって文章推敲を促す．

3.4.4 文章のスコア表示

表示例を図 2 の④に示す．文章の内容を点数化，また可視化表示することで，現在の文章を定量的に評価する．3.4.1 節で抽出した好意と悪意を表す単語の含まれる文の数，また，好意と悪意の表現を含む文の数を受けて，全文中における割合をパーセンテージで算出し，それぞれ薄い赤色と薄い青色の棒グラフで可視化する．好意と悪意を表す棒グラフの原点はそれぞれ左右の端であり，100 %の点は棒グラフの中央である．また，好意を表す単語が含まれる，または好意の表現を含む文の割合が 50 %の時に 100 点になるようにスコアを算出し表示している．文章中の色付きの単語やハイライトしている文の割合を視覚的に表現して，ユーザの文章推敲を促す．

3.4.5 文章の編集機能

この機能は以下の 3 つの機能に分かれている．

- 文章を編集する機能．図 2 の①の領域で行うことができ，文章を 1 文ずつ編集することが出来る．文末に全角の「.」または「。」を追加することで文の数を増やすことが出来る．
- 悪意を表す単語や文の推敲を促すチェックボックス機能．この機能は図 2 の⑤の領域で行うことが出来る．悪意を表す単語が含まれる文，または悪意の表現を含む文の先頭のチェックボックスに

チェックをすることで，チェックされた文を悪意を表す単語，また文として扱わず，図 2 の④の棒グラフを再描写する．悪意を表す単語として表示された単語，または文について，推敲しなくてもそのままの表現で問題ない場合にチェックボックスにチェックをすることで，問題ないことを確認する．文章推敲を促すメッセージ表示機能との連携によって，ユーザに悪意を表す単語や文の確認を促すことが出来る．

- システムの再描写機能．この機能は図 2 の⑥の領域で行うことが出来る．編集した文章をシステムに再入力するボタンである．再入力する際に，チェックボックスへのチェックがリセットされるため，編集内容に応じた再確認を促すことが出来る．

これらの機能によってユーザに文章推敲を促し，読み手に好感持たれる文章の推敲を支援する．

4 文章推敲支援機能の効果の検証

4.1 実験目的

本実験では，文章推敲支援インタフェースにおける深層学習を用いた好意と悪意の表現を含む文の可視化を利用することによって，文章中の読み手に好感を持ってもらえる文章の推敲を支援出来るかを評価する．具体的には，悪意を含む表現を文章中から削除し，好意を含む表現を文章中により多く追加することで，読み手の印象を良くするという文章推敲の支援目標を達成できるかを評価する実験を行った．

この実験目的を達成するために，以下の 2 種類の実験を行なった．1 つは，4.2 節で述べる文章推敲支援システムを用いた文章作成実験，もう 1 つは，4.3 節で述べる推敲された文章の内容評価実験となる．それぞれに別の被験者を用意し，実験を行なった．

4.2 文章推敲支援システムを用いた文章作成実験

4.2.1 文章推敲支援システムを用いた文章作成実験の実験方法

表 4 に提示する 4 つの状況において，指定の目的を達成するメールを，独力で作成してもらった．ここで，状況 1，3 は書き手の悪意が表れやすい状況，状況 2，4 は書き手の好意が表れやすい状況として用意した．それぞれの状況において，システムを使用する事で，書き手の悪意を表す表現が削除され，好意を表す表現が多く追加されることを期待している．次に，作成したメールを「悪意を含む表現を使わない，また好意を含む表現を増やしたメールに修正してください」という指示のもと，独力で推敲してもらった．その後，独力で推敲したメールを提案システムに入力してもらい，システムを用いて同様の指示のもと，再度メールの推敲を行ってもらった．これを 11 名の理系学生を被験者として行った．ここで，独力で推敲したメールを「システム使用前のメール」とし，システムを用いて推敲した後のメールを「システム使用後のメール」と呼ぶ．

表 4: メールの作成課題の状況

	送信相手	送信相手について	送信したいメール
1	同じ サークルの 同性の先輩	自己中心的な性格で 普段からあなたを含め 周りの人のことを全く 考えない発言や行動を する先輩.	あなたが企画したサークル 旅行について先輩が楽しめる 企画が少ないから計画を大幅 に変更するようにと自分勝手 な事を言われた．あなたの 周りのメンバーは，計画に 概ね同意している．そこで 先輩に自分勝手な主張は やめてもらい普段から皆 困っていると文句を言って 欲しいとメンバーから 言われている．
2	同じ学科の 異性の先輩	成績優秀でいつも やさしく，先生や周り の人からの評価も高く あなた自身も憧れて いる先輩.	期末試験に向けて，勉強で わからないところを 質問したら個人的に時間を 取ってくれて，丁寧に わかりやすく教えてもらった ので感謝のお礼を伝えたい． また先輩ともっと仲良く なりたい．
3	同じ 研究室の 同性の後輩	自分を含め，後輩から 見て先輩にあたる人達 に，普段からタメ口で 文句を言うなど， 研究室の多くの人から その態度や言葉遣いが 問題視されている．	研究室の研究報告会で， 質問者である自分がした指摘 に対して不満そうな顔をした 上で，ぶっきらぼうに「それ はない」と言った．先生から はその指摘は正しいという 言葉も貰っている．報告会 での態度を謝るように，また 普段からも最低限の礼儀を わきまえるよう指導したい．
4	同じ サークルの 気になる 異性	普段から仲良く話を する相手．とても気が 合って数人で一緒に 遊びに行く事もある． 服装にこだわる おしゃれさん.	次の土曜日に，普段の おしゃれさから自分の服装に もアドバイスをもらいたい という話から入り自分と二人 で遊びに行くように 誘いたい．

4.2.2 文章推敲支援システムを用いた文章作成実験の結果と考察

まず，文章の修正がシステムのどの表示による修正か確認するために，システム使用前後で修正が行われた文について，その修正方法を以下のように区別する．好意と悪意を表す単語を追加，削除することでの修正を「単語単位での修正」と呼ぶ．また，好意と悪意を表す単語の追加，削除以外での文単位での修正を「文単位での修正」と呼ぶ．文単位での修正は，3.4.1 項で

前述した深層学習による好意や悪意の表現を含む文のハイライト表示を用いて修正された可能性が高いと考えられる．単語単位での修正，文単位での修正どちらも行われた場合は「単語と文による修正」と呼ぶ．

各被験者がシステムを用いて修正した文について，その修正方法の内訳を確認するために，各状況毎のシステム使用前後で修正された文の内訳を図 4 に示す．図 4 からわかるように，状況 1，3，4 ではシステムを用いることで全文中半数以上が修正がされており，状況 2 においても 27 %程度の文が修正されている．また，修正の内訳のうち，文単位の修正，単語と文での修正が大多数を占めている．これにより，システムの深層学習による好意や悪意の表現を含む文のハイライト表示を用いて文章のおそよ半数程度の文修正するように文章推敲を促すことができたと考えられる．

また，修正が行われた文のうち，修正によって深層学習の結果によるハイライト表示がどのように変化したか確認するため，各状況毎のシステム使用前後での修正によるハイライト表示の変化の内訳を図 5 に示す．ここで，「修正でハイライトが改善しなかった文の数」とは，修正された文のうち，修正によって新たに好意ハイライトが表示された文，悪意ハイライトが消えた文以外の文を指す．図 5 より，各状況において，修正された文のうち，およそ半数以上の文で新たに好意のハイライトが表示されるか悪意のハイライトが消えている．特に，悪意の表れやすい状況である状況 1，状況 3 では修正によって悪意のハイライトが多く消えている．また，好意の表れやすい状況である状況 2 で好意のハイライトが多く追加されている．これより，深層学習による好意や悪意の表現を含む文のハイライト表示を用いて好意のハイライトが表示される文を増やし，悪意のハイライトが表示される文を減らすように文章推敲を促すことができたと考えられる．

4.3 推敲された文章の内容評価実験

4.3.1 推敲された文章の内容評価実験の実験方法

4.2 節の実験によって推敲されたメールが，自分宛てに送られてきたという仮定で，メールの送り主にどの程度好感を持ったか，理由とともに 7 点満点で評価してもらう．評価点と評価基準を表 5 に示す．また，システム使用前とシステム使用後のメールの各文を，どちらがシステム使用後か明示せずに，どちらに好感を持ったかを評価してもらう．これを 4.2 節の実験で作成された 11 名 (被験者数) × 2 種類 (システムなしでの推敲とシステム有りでの推敲) × 4 種類 (状況) = 88 種類 のメールに対して行ってもらった．この実験を 8 名の理系大学生を評価者として行った．

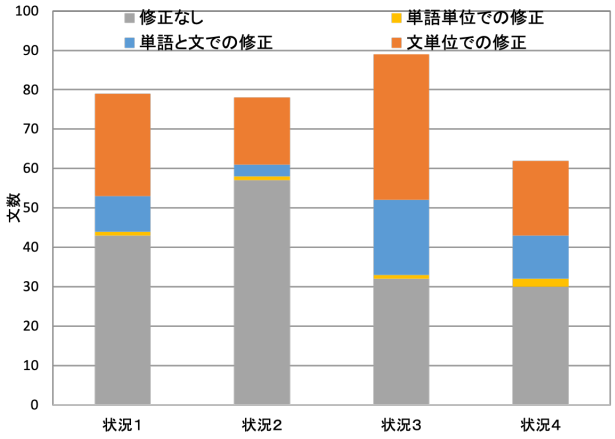

図 4: 各状況毎のシステム使用前後で修正された文の内訳

表 5: メールの評価点と評価基準

点数	評価
7	メールの送り主に，強く好感が持てる．
6	メールの送り主に，割りと交換が持てる．
5	メールの送り主に，少し好感が持てる．
4	メールの送り主に，特になんとも思わない．
3	メールの送り主に，少しだけ好感が持てない．
2	メールの送り主に，あまり好感が持てない．
1	メールの送り主に，全く好感が持てない．

4.3.2 推敲された文章の内容評価実験の結果と考察

システム使用前後で読み手により好感を持ってもらえる文章に修正されたか検証するために，各状況におけるシステム使用前と使用後のメールの評価点の被験者毎の平均を図 6 に示す．ここでの評価点の差とは，システム使用後のメールの評価点から使用前のメールの評価点を引いたものとする．本実験では，既に一度推敲したメールを，システムによってより読み手に好感を持たれるメールに修正していることに注意されたい．

図 6 より，ほとんどの被験者がシステム使用前に比べてシステム使用後の評価点が高いことがわかる．特に，状況 1，状況 3 の被験者毎の評価点の差の平均はそれぞれ 1.0，0.9 となっている．これは，状況 1，状況 3 がそれぞれ「先輩に文句を言う」状況と「後輩に注意する」状況であるため，深層学習によって，明示的に悪意を表す単語が使われないが，悪意を表す文をハイライトすることで文章修正が行われたためだと考えられる．これより，悪意を含む表現が使われやすい状況においては，システムを用いる事でおよそ 1 段階程度の評価をあげる程度の文章の推敲を支援できることがわかる．しかし，状況 2，状況 4 に関しては評価点の差があまり大きくなかった．ここで，状況 2，状

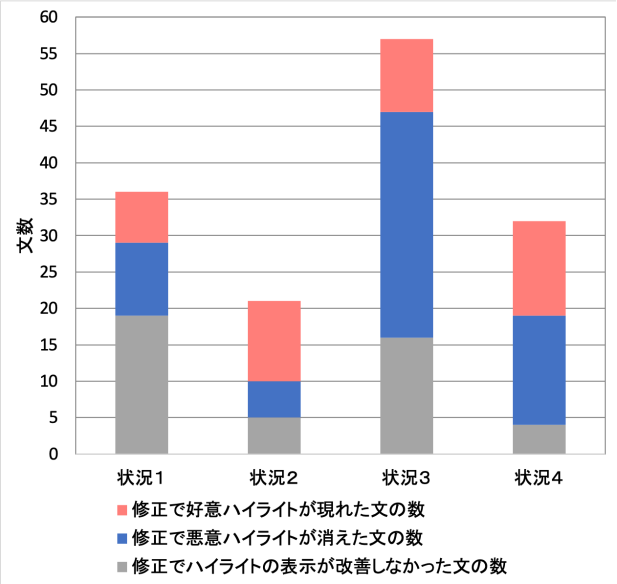

図 5: 各状況毎のシステム使用前後での修正によるハイライト表示の変化の内訳

況 4 はそれぞれ「先輩にお礼を言う」状況と「異性をデートに誘う」状況となっており，共に好意を含む表現が表れやすい状況となる．このため，システム使用前の評価点が悪意を含む表現が表れやすい状況に比べて 1 点ほど高く，独力でもある程度好意的な文になっていたと考えられる．全体として，評価はシステム使用前後で上昇した．特に，悪意を含む表現が発生しやすい状況では，深層学習を用いた好意と悪意の表現を含む文の可視化に意味があるとわかった．

システム使用前後で評価点が大きく上がったまたは下がったメールについて，システム使用前後のメールを図 7，図 8 に示す．ここで，図 7，図 8 ではシステム使用前後での差異を赤文字で表示している．また，深層学習の結果によるハイライト表示で好意のハイライトがされた文を薄い赤色で，悪意のハイライトがされた文を薄い青色で塗りつぶしている．

図 7 は状況 1 における被験者 E のシステム使用前後のメールと好意悪意度合いの差を表している．被験者 E は，悪意ハイライトのされた「苦情が出ています」という文を修正して「相談をいくつか受けています」という表現にすることで，悪意ハイライトを消し，より読み手に悪意を与えない文にした．また,「ご理解いただければと思います」という表現を「ご理解いただければ幸いです」という表現に修正し，好意ハイライトを追加することで，より柔らかい印象を受ける文にした．この修正によって評価を 4.0 点から 5.4 点にあげている．

図 8 は状況 1 における被験者 A のシステム使用前後

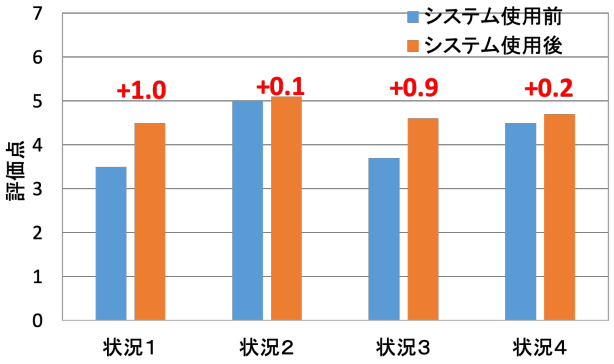

図 6: 各状況ごとのシステム使用前後での評価点の平均と差

のメールと好意悪意度合いの差を表している．被験者 A は，修正前後でほとんど修正をしておらず，好意を含む文も追加していない．語尾に「ね」をつけた修正を行っているが，これによって上から目線の文章になっており，評価点は 4.6 点から 4.0 点に下がっている．

全体として，評価点が大きく上昇したメールでは，悪意を含む表現を無くし好意を含む表現を追加する修正が多かった．これにより，システムを用いて悪意を含む表現を無くし好意を含む表現を追加する事で，読み手に好感を持たれる文章を作成できると考えられる．また，修正を促された時に，どのように修正したらよいかわからない場合もあると考えられる．

5 おわりに

読み手に好感持たれる文章推敲支援を行うために，深層学習を用いて文章の書き手の読み手に対する好意，悪意の表現を含む文を抽出して可視化し，書き手に文章推敲を促すシステムを実装した．

また，実装した文章推敲支援システムが，読み手に好感を持たれる文章の推敲に有用であるかを検証するために評価実験を行った．ある状況において作成したメールを，一度独力で推敲したうえで文章推敲支援システムを利用して再度推敲してもらい，システム使用前後の推敲されたメールを別の被験者に評価してもらった．この実験より，文章推敲支援システムを用いて文章推敲する事で，より読み手に好感持たれる文章を作成できることがわかった．

今後の展望としては，修正が望まれる箇所に対して，修正案を提示できるシステムへの改良が期待される．

システム使用前(評価点4.0)	システム使用後(評価点5.4)
○○先輩、いつもお世話になっております、XXです。	○○先輩、いつもお世話になっております、XXです。
(中略)	(中略)
また、この場をお借りして申し上げますが、以前からサークル内で先輩の振る舞いに対し少し苦情が出ています。	また、この場をお借りして申し上げますが、以前からサークル内での先輩の振る舞いについて相談をいくつか受けています。
(中略)	(中略)
重ね重ね失礼なことを述べましたが、どうかご理解いただければと思います。	重ね重ね失礼なことを述べましたが、どうかご理解いただければ幸いです。
以上、どうかよろしくお願いします。	以上、どうかよろしくお願いします。

図 7: 各状況 1 における被験者 E のシステム使用前後のメール (評価点 1.4 上昇)

システム使用前(評価点4.6)	システム使用後(評価点4.0)
○○様, お世話になっております, ××です.	○○様, お世話になっております, ××です.
先日, ご連絡いただいた企画変更につきまして, 部内で検討した結果, 変更せずに企画を行う事となりました.	先日, ご連絡いただいた企画変更につきまして, 部内で検討した結果, 変更せずに企画を行う事となりました.
大変申し訳ありませんが, ご理解の程よろしくお願いいたします.	大変申し訳ありませんが, ご理解の程よろしくお願いいたします.
次回, 企画するときは一緒に考えましょう！.	次回, 企画するときは一緒に考えましょうね！.
よろしくお願いします.	よろしくお願いしますね.

図 8: 各状況 1 における被験者 A のシステム使用前後のメール (評価点 0.6 下降)

謝辞

本研究における深層学習を用いた好意と悪意の表現を含む文の抽出のために株式会社ホットリンクより, 大規模日本語 SNS コーパスによる文分散表現モデル hottoSNS-w2v[9] を利用させて頂いたことついて感謝する.

参考文献

[1] 総務省 : ネット依存など新たな課題とインターネットリテラシーの重要性, 総務省平成 26 年版情報通信白書, pp.283–303 (2014)

[2] 庵 翔太, 砂山 渡, 畑中 裕司, 小郷原 一智 : 良好な人間関係構築のための好意と悪意を表す単語の可視化による文章作成支援, 第 32 回人工知能学会全国大会論文集, 3F2-OS-12b-03 (2018)

[3] 板倉 由知, 白井 治彦, 黒岩 丈介, 小高 知宏, 小倉 久和 : 様々な文書を対象とした段落一貫性の解析, 情報処理学会研究報告自然言語処理（NL）, Vol.192, No.9, pp.1–6 (2009)

[4] 大野 博之, 稲積 宏誠 : 文の接続関係を利用した論理方向の可視化による技術文章作成支援, 電子情報通信学会技術研究報告, Vol.107, No.48, pp.27–32 (2015)

[5] 徳岡 拓弥, 竹内 章, 國近 秀信 : 変更履歴を用いた英語文章作成支援システムの実現, 第 30 回人工知能学会全国大会論文集, 1C3-1 (2016)

[6] 石川 葉子, 水上 雅博, 吉野 幸一郎, Sakti Sakriani, 鈴木 優, 中村 哲 : 感情表現を用いた説得対話システム, 人工知能学会論文誌, Vol.33, No.1, pp.1–9 (2018)

[7] 長谷川 貴之, 鍛治 伸裕, 吉永 直樹, 豊田 正史 : オンライン上の対話における聞き手の感情予測と喚起, 人工知能学会論文誌, Vol.29, No.1, pp.90–99 (2014)

[8] Twitter.com, (URL)https://dev.twitter.com/docs/api/

[9] 日本語大規模 SNS+Web コーパスによる単語分散表現モデルの公開 : hottoSNS-w2v の配布, (URL)https://www.hottolink.co.jp/blog/20190311_101674/

[10] Total Environment for Text Data Mining（テキストデータマイニングのための統合環境）, (URL)https://tetdm.jp

[11] 砂山 渡, 高間 康史, 西原 陽子, 徳永 秀和, 串間 宗夫, 阿部 秀尚, 梶並 知記 : テキストデータマイニングのための統合環境 TETDM の開発, 人工知能学会論文誌, Vol.28, No.1, pp.1–12 (2013)