卒業論文

周辺語を用いた感情語が表す多様な感情の推定手法

吉山 隆聖

2022 年 2 月 9 日

岐阜大学 工学部 電気電子・情報工学科 情報コース
鈴木研究室

本論文は岐阜大学工学部に
学士（工学）授与の要件として提出した卒業論文である.

吉山 隆聖

指導教員：
鈴木 優 准教授

周辺語を用いた感情語が表す多様な感情の推定手法*

吉山 隆聖

内容梗概

本研究では，同じ感情語が表す多様な感情の推定を自動的に判定することを目的とする．絵文字のような役割で，用いられている感情語が存在する．しかし，同じ感情語が用いられているにも関わらず，異なる感情を表現するために用いられていることがある．そのため，受け手によって異なる捉え方をする事も多くある．そこで，機械学習を用いることによって，同じ感情語が用いられる文章がどのような感情を持っているかを判定する．感情の定義を行い，Twitter のツイートを収集し，クラウドソーシングを用いてデータセットの作成を行った．この作成したデータセットを用いて，ファインチューニングした BERT を用いて，分類機を作成する．この分類機を用いて，同じ感情語を持つ文章がどのような感情を持つかを判定する．この分類性能に関する評価実験を行った．

キーワード

感情分析，機械学習，テキスト分類

*岐阜大学 工学部 電気電子・情報工学科 情報コース 卒業論文, 学籍番号: 1183033169, 2022 年 2 月 9 日.

目次

図目次

表目次

第 1 章　はじめに

インターネット上でのメッセージのやり取りは，非常に増えている．そのやり取りの中で感情を適切に伝えるためや感情を強調するための手段として，絵文字や顔文字がある．これらは，非常に多くの種類があり細かな感情を伝えるのに用いられている．また,「笑」というような漢字も絵文字や顔文字と同じように感情を伝えるためや強調するために，文章に付け加えられている場合もある．本論文では，この「笑」という文字に焦点を当てる．

笑という文字を見ると，微笑みや笑顔のような幸福感や楽しさなどのポジティブな感情を示す言葉や，何かしら面白かったことを伝えるような場面を連想する人が多いと考える．一方で，嘲笑や冷笑などの嫌悪や軽蔑などのネガティブな感情を示す言葉や自虐的な発言で笑いを誘うような場面も考えられる．例えば，友達と久しぶりに会った笑，というテキストでは幸福感や楽しさのようなポジティブな感情を表現していると考えられる．スマホ壊れた笑，というテキストでは悲しみのようなネガティブな感情を表現していると考えられる．また，眠い笑，というテキストでは特に感情が感じられないような場合も用いられている．このように，テキスト上で絵文字や顔文字のように現れる「笑」という文字にはポジティブな感情やネガティブな感情，特に感情が込められていない場合でも用いられており，感情を伝えたり，強調するものとしては曖昧なものになっている．同じ感情語であるにも関わらず多様な感情表現に用いられていることが問題であると考える．

そこで，本研究では，その感情を自動的に明確にするためのシステムを作るために，機械学習を用いることを提案する．明確にすることができれば，感情の捉え方によるトラブルを防ぐことや，不快だと感じる感情を避ける事も可能であると考える．そのために，本研究では BERT を用いた．様々な自然言語処理タスクにおいて汎用性の高いモデルである BERT による分類器を作成する．この分類機を用いることによって，感情を基準にしたブロック機能・フィルター機能や，表現をより分かりやすいものに変換するシステムを作る際に役立つと考える．例えば，あるユーザーが不快感を感じるため，攻撃的な感情を持つテキストを目にしたくないとする．分類機をフィルターの役割として用いることによって，目にする機会を無くすことに役立つと考える．また，笑のように多様な感情を表現するために同じ感情

語が用いられている場合，どの感情として捉えればよいか判断に迷う場合がある．これを分類機によって判定し，絵文字や顔文字のような種類が多く，感情が捉えやすいものに置き換えるというようなシステムを作る際に用いることができると考える．どちらの場合も，人手を介すことなく自動で判断できる部分もメリットがあると考える．

本研究に着手するにあたって，笑を含む文章がどのような感情が存在しているかの定義と笑を含むテキストのデータセットの作成を行った．まず，笑を含むテキストデータに関しては，Twitter API を用いて収集した．どのような感情が存在しているかの定義に関しては，私がその収集したテキストデータやプルチックの感情輪やエクマンの定義した基本感情，心理学の図書などを参考にした．その結果，ポジティブ，ネガティブ，ニュートラル，ポジティブ＋ネガティブの 4 つを大枠として定義し，ネガティブの中でのみ攻撃的，非攻撃的，自虐の 3 つを定義した．そのため，付与する感情ラベルはポジティブ，ニュートラル，ポジティブ＋ネガティブ，攻撃的，非攻撃的，自虐の 6 つである．データセットの作成は，クラウドソーシングを用いて行った．収集したツイーと一つに対して五人に評価してもらい，過半数が同じ感情ラベルに割り振ったものをそのテキストに表現されている感情としてラベルを付与した．その結果，ニュートラルが 11772，ポジティブが 8453，ポジティブ＋ネガティブが 906，攻撃的が 958，非攻撃的が 1686，自虐が 850，の合計 24，625 のデータを収集することができた．

作成したデータセットを用いて，BERT を用いて学習させた．学習はニュートラル，ネガティブ，ポジティブの 3 つの大枠の感情で分類する場合とニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ネガティブ＋ポジティブの私が定めた 6 つの感情で提案手法の評価実験として，評価値に基づくモデルの分類性能の評価を行った．また，データの偏りを無くすためにデータの水増しを行う前後での分類性能の評価も行った．大枠での分類に関しては，単純な正解率である Accuracy だけに注目した場合，ランダムな予測を行う場合よりも，高い数値を取ることが確認できた．また，データ数の偏りを水増しを行うことで多少の効果があることがわかった．6 つの感情での分類に関しては，単純な正解率である Accuracy だけに注目した場合，こちらもランダムな予測を行う場合よりも，高い数値を取ることが確認できた．また，水増しを行った後の攻撃的な感情を持つと判断されたテキストに関し

ては，元々のデータ数が少ない中ではかなり高い評価値を取ることがわかった．しかし，データの偏りの解決策として水増しを行うことが効果があるとは言えない結果となった．収集したデータの中でも，データ数が少ないラベルのデータが特徴を捉えるには不足していることが問題点として挙げられる．

本論文における貢献は以下の通りである．

- 同じ感情語が用いらており，異なる感情を持つ文章は機械学習で分類可能である．

本論文の構成は以下のとおりである．2 章では本論文にて用いた技術や手法についての基本的事項を述べる．3 章では関連研究について述べる．4 章では本論文の提案手法について述べる．5 章では私が定めたラベルの定義について述べる．6 章では評価実験の目的と内容，結果と考察について述べる．7 章では本論文のまとめと今後の課題について述べる．

第 2 章　基本的事項

本論文にて用いた技術や手法について，基本事項を述べる．

2.1　Twitter API

Twitter API とは，ツイートの取得・投稿，いいねやリツイートなど Twitter のサービスを公式のウェブサイトを経由することなく直接利用できるサービスのこと．ユーザーをフォローすることやツイートの検索も行うことが可能．利用するには API キーを取得する必要がある．

2.2　クラウドソーシング

クラウドソーシングとは，インターネット上で企業や個人事業主などが不特定多数の人々に業務を発注する業務形態のこと．crowd(大衆) と sourcing(調達) を組み合わせた造語である．特定の業者や個人に業務を委託するアウトソーシングとは異なり，不特定多数の業者や個人に依頼するという特徴がある．そのため，発注側は人材を採用するコストを抑え必要な時にピンポイントで業務を発注でき，受注側は自分のスキルを活かしながら，好きなタイミングで働くことができるため，副業として働くことができるというメリットがある．しかし，直接コミュニケーションをとる機会が少ないため，行き違いでトラブルになる可能性がある．

クラウドソーシングを行う場合，クラウドワークスなどの仲介サイトを通して，作業を依頼することができる．

2.3　BERT

BERT(Bidirectional Encoder Representations from Transformers) とは，Transformer の Encoder を使用した 12 層のニューラルネットワークモデルである．Jakob Devlin らの論文 [1] で発表された．モデルの構造を修正せずとも，転移学習することで，様々な自然言語処理タスクに応用できる汎用性の高いモデルと

なっている．事前学習として，複数の穴がある文章に対して穴埋め単語を予測するMLM(Masked Language Modeling) というタスクと入力された二文が連続している文であるかどうかを NSP(Next Sentence Prediction) を長い文章を含むデータセットを用いて行っている．事前学習で用いるデータには，ラベルは付与されていない．MLM では，入力の 15% を確率的に [MASK] トークンに置き換えを行い，この内の 80% を [MASK] トークンに置換，10% をランダムなトークンに置換，10% は元のまま変えずに行い，置換前のトークンは何であるかを予測する穴埋め問題を行う．NSP では，50% を二文が関係するもの，50% を二文が関係しないものとして，片方の文に続いてもう片方の文の意味が通るかどうかを予測する．この 2 つを同時に行うことで双方向学習を実現している．そのため，文脈を読むことが可能となった．事前学習で得たネットワークの重みをファインチューニングをすることで，他のタスクにおいても高い精度を期待できる．ファインチューニングとは，学習済みモデルの重みの一部を再学習させる手法である．これを行うことにより，目的に合ったモデルを作ることができる．またファインチューニングを行う際は，ラベル付きのデータが必要となる．

2.4 評価指標

本研究では，モデルの評価を行うために，Accuracy, Recall, Precision, F 値を用いる．Accuracy は予測したモデルの単純な正解率を見ることのできる指標である．Accuracy のみの評価では，本研究のデータのように偏りが生じているデータを扱う場合に，すべての予測を偏っているデータに予測してしまえば自ずと Accuracy は高い値を取ってしまうが，他の予測がされないため良い予測であるとは言うことができない．例えば，正が 5 件，負が 95 件のデータがあり，すべてを負と予測しても，95% は正解していることになる．

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{2.4.1}$$

Precision は適合率とも呼ばれ，正と予測したものがどれだけ正しかったかを見ることのできる指標である．確実なもののみ正であると予測し，判断に迷うものは

負と予測すると高い値を取る．Precision は以下の式で求められる．

$$Precision = \frac{TP}{TP + FP} \tag{2.4.2}$$

Recall は再現率とも呼ばれ，実際に正であるデータのうち，どれだけのデータが正であると予測されたかを見ることのできる指標である．予測するデータをすべて正であると予測すれと高い値を取る．Recall は以下の式で求められる．

$$Recall = \frac{TP}{TP + FN} \tag{2.4.3}$$

F 値は Recall と Precision の調和平均を取ることで求めることができる．Precision と Recall はどちらかが高くなると，もう一方は低くなるというトレードオフの関係にある．F 値は Recall と Precision の両方の値を考慮した評価をすることができる．F 値は以下の式で求められる．

$$F\text{ 値} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2Recall \cdot Precision}{Recall + Precision} \tag{2.4.4}$$

表 2.1　評価値導出のための混同行列

	予測: 正	予測: 負
実際: 正	TP	FN
実際: 負	FP	TN

TP　正と予測されたうち，実際も正であるデータの数.
FP　正と予測されたうち，実際が負であるデータの数.
FN　負と予測されたうち，実際が正であるデータの数.
TN　負と予測されたうち，実際も負であるデータの数.

第 3 章　関連研究

感情を伝えたり，強調するために用いられる絵文字の研究や顔文字の研究は存在する．また，テキストの感情分析を行っている研究も存在する．

山本らの研究 [2] では，絵文字の曖昧性解消に対しての手法が提案されている．絵文字を内容語，内容添加，モダリティ，装飾の 4 つの役割に分ける．共起語に注目して，絵文字の持っている語義を抽出する．それらの語義との類似度，絵文字の表記，絵文字前後の品詞，同一絵文字の有無の 4 つの素性を用いて，役割の推定を行う．本研究と異なる部分は，モダリティに着目しその部分を掘り下げている部分である．

黒崎らの研究 [3] では，顔文字の感情分類に対しての手法が提案されている．Word2Vec を用いることで単語・顔文字のベクトル表現を教師なし学習で取得することができる．この取得した顔文字のベクトルと感情を表す語との距離のをそれぞれ測定し，最も距離が近くなった感情をその顔文字が持つ感情として推定している．

源田らの研究 [4] では，黒崎らの研究 [3] を先行研究とし，1 つの顔文字で Word2Vec を用いたベクトルの生成を行うのではなく，構成要素ごとの意味を分析し，顔文字全体が表現する感情の分析を行う．こちらの研究も単語との距離を測定し，最も距離が近くなった感情をその顔文字が持つ感情として推定している．

これらの研究と本研究の似ている部分は，教師の有無はあるが感情を伝える，強調するための役割を持つものの感情推定を行おうとしている部分である．これらの研究と本研究の違いは，多くの種類やパターンが存在している顔文字をいくつかの感情に推定するのではなく，1 つの表現のみを扱いいくつかの感情に分類を試みているという点がある．また，BERT を用いるという部分でも異なる．

大町らの研究 [5] では，顔文字とオノマトペが用いられている文章を対象として，両者の複合的要素から抽出される感情成分に着目し柔軟な感情推定を行うことを目標としている．Twitter のツイートを対象とし，顔文字とオノマトペ，それぞれに表現される感情と組み合わせに着目し，2 つが混合された形である混合形がどのような感情を表しているか，また，文章と混合形が持つ感情の組み合わせによっても，どのような感情が表されているかを述べている．本研究と似ている部分は，感情を伝達するための役割を持つものに着目している点である．しかし，実際に扱うもの

は，顔文字とオノマトペであり，本研究では，笑という部分が異なる．

吉田らの研究 [6] では，Twitter のツイートから顔文字を自動的に抽出し，そのつぶやきから推定した感情を付与したデータベースの構築の手法を提案している．感情の推定には感情表現辞典を用いており，感情語に対応する感情を顔文字が持つ感情として推定を行っている．感情の推定をあらかじめ決められている語ではなく，文脈を考慮できる BERT による推定を行う部分で異なる．

堀宮らの研究 [7] では，ツイートのリプライを利用したリプライ元ツイートの感情推定を行っている．Ekman の定義した基本 6 感情に無感情を加えた 7 ラベルのいずれか 1 つを付与し，TF-IDF で重みを算出した後，SVM で学習させ分類機を作成し，感情を判定させている．また，Ekman の定義した基本 6 感情である幸福感，驚き，恐れ，悲しみ，怒り，嫌悪の 6 感情を扱い，各クラスの感情であるかそうでないかの 2 クラス分類として扱っている．本研究と似ている部分は，リプライを扱っているかどうかの違いはあるが，Twitter を用いたデータセットを作成・利用している部分である．本研究と異なる部分は，使用するモデルが異なる部分，2 クラス分類としている．本研究では SVM ではなく，BERT を用いており，多クラス分類として捉えている．

若井らの研究 [8] では，顔文字だけでなく Twitter 特有表現に着目して，映画の実況ツイートに対しての感情分析を行っている．Twitter 特有表現とは，同じ文字を繰り返し叫んでいるような表現をしているもの，としている．この Twitter 特有表現の有無でユーザー実験を行い，感情の強弱に影響を及ぼすかの実験を行っている．本研究と似ている部分は，SNS 特有の表現・記法を着目しているところである．笑という文字も SNS 特有の表現である．本研究と異なる部分は，機械学習を用いている部分である．

松林らの研究 [9] では，Twitter のツイートを用いて感情分析を行い，感情の推移の可視化やいいね数などが一定の値を超えると通知が送られるなどの機能を実装した Twtter モニタリングシステムに適用している．感情分析には Wikipedia のコーパスで学習した Word2Vec で特徴ベクトルを取得し，ランダムフォレストを用いて分類を行っている．

中澤らの研究 [10] では，BERT を用いた単文の極性推定手法の提案をしている．データセットとして読売新聞のニュース記事を扱い，BERT モデルを用いて感情極

性推定を行っている．本研究と似ている部分は，BERT モデルを用いているところである．本研究と異なる部分は感情の定義である．感情極性値とは，単語ごとに設定されている値を指し，最大値を 1，最小値を-1 の値を取る．また，値は 1 に近づくほどポジティブな感情をもつ単語であり，-1 に近づくほどネガティブな感情をもつ単語であることを示す．本研究では，ポジティブとネガティブの間ではなく，笑が付与されているテキストには 6 つの感情が存在するとして定義している．

高津らの研究 [11] では，ニュース記事を対象にし，BERT と BiLSTM-CRF を組み合わせたモデルを用いて，タイトルと文の感情ラベルを推定する手法を提案している．感情ラベルはポジティブ，ニュートラル，ネガティブの 3 つである．本研究と異なる部分は，感情ラベルの数である．本研究では，ポジティブ，ニュートラルに加え，ネガティブを攻撃的，非攻撃的，自虐と細分化して扱い，ポジティブ＋ネガティブのラベルを加えた 6 ラベルを定義し，扱っている．

第 4 章　提案手法

本論文では，笑を含む文章がどのような感情を持つかを表す感情分類問題としても捉えることにする．ラベル付けされたテキストを用いて BERT のファインチューニングを行う．このファインチューニングを行った分類機を用いて，笑を含むテキストが機械学習で自動的に分類可能であるかを行う．提案手法の流れを図 4.1 に示す.

4.1　データセットの作成

まず，TwitterAPI を用いてツイートを収集する．次に，収集したツイートに対して出会い系ツイートや bot 等の望まないデータをできる限り取り除いた処理を施した．その後，クラウドソーシングを用いて，ラベル付けを行ってもらった．１つのツイートに対して 5 人の作業者に判断してもらいどの感情ラベルに適切であるかを判定してもらった．また，指示は，各ラベルの説明と例文を提示して行った．感情ラベルとはニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ポジティブ＋ネガティブの 6 つのラベルに望まないデータを取り除く段階で漏れたデータを排除するためのその他，また，ネガティブの中で判断がしにくいものに対して判定不可の 2 つを加えた 8 つの選択肢から選択する形となっている.

3 人以上の作業者が同じラベルを選択した場合，そのラベルを感情ラベルとして採用する.

4.2　モデル構築・学習

BERT は多様な自然言語処理タスクにおいて，モデルの構造を修正することなく，転移学習のみで様々なタスクに応用でき，高い精度を出すことのできる，汎用性の高いモデルである.

本論文では，訓練済みの BERT モデルである，東北大学の乾・鈴木研究室の訓

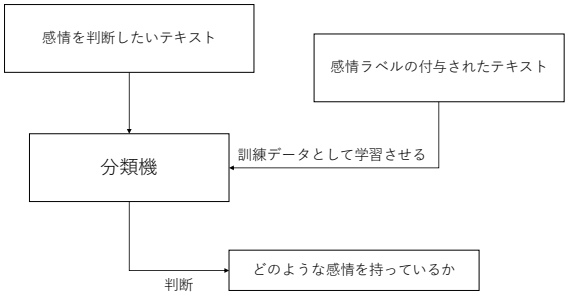

図 4.1　提案手法の概要

練済み日本語 BERT モデル*を使用した．この訓練済みモデルは日本語 Wikipedia で事前学習が行われており，語彙数は 32，000 となっている．また，この BERT モデルは 12 層である．これに出力層を 1 層加えた 13 層のモデルを構築した．4.1 節に従って作成したデータセットを用いてファインチューニングを行うことで，笑を含むテキストの分類が可能になると考えた．その際，重みの更新は最終層だけを行う．しかし，本研究で取り扱いたい笑と笑った，爆笑したなどのような動詞などを活用した形，もしくは，熟語などの一部になっているものとの区別が形態素解析をする際に区別できなくなってしまうという問題点がある．そのため，前処理として，本論文で取り扱いたい笑に関しては，「*?*」という記号に置き換えることを行う．例えば，爆笑した笑，というテキストは，爆笑した*?*，というテキストになった状態で入力され，形態素解析される．単語列をベクトルに変換する Embedding 層に新たに「*?*」を語彙として追加し，学習する際に区別できるようにしている．そのため，語彙数は 32,001 となっている．

分類を行う流れを示す．6 クラス分類を行う場合を考える．まず，笑を含むテキストを形態素解析して，単語ごとに BERT モデルに対応した ID 化を行い数値化する．形態素解析には mecab-ipadic-NEologd を辞書とした，MeCab を用いて行う．単語の ID 化は BERT の訓練済みモデルの tokenizer で行う．入力文章を ID 化し，

*https://github.com/cl-tohoku/bert-japanese

単語列にしたものを構築したモデルに入力する．その後，単語を Embedding 層にて，事前学習の際に得られた 768 次元の分散表現を利用した単語ベクトルに変換する．その後，BERT を用いた，12 層＋出力層の合計 13 層の構築したモデルで計算を行う．計算の結果，出力として 6 次元のベクトルを得ることができる．その出力された 6 次元のベクトルを Softmax 関数を用いて各要素の値を 0 ～ 1 の間で表すことができる．また，この 6 次元の要素の和は 1 になる．Softmax 関数は以下のように定義される．

$$Softmax(x_i) = \frac{\exp(x_i)}{\exp(x_1) + \exp(x_2) + \cdots + \exp(x_n)} \tag{4.2.1}$$

Softmax 関数を通して計算された 6 次元のベクトルの各要素は，そのままどのラベルであるかの確率を表現している．よって，入力された笑を含むテキストに関して，出力の確率の最大値を取り，ラベルの予測を行う．

第 5 章　ラベルの定義

感情の定義はさまざまである．表情への感情表出を研究しているポール・エクマンは幸せ，驚き，悲しみ，恐れ，嫌悪，怒りの 6 つを基本感情とした．プルチックは喜び，信頼，驚き，期待，恐れ，悲しみ，嫌悪，怒りの 8 つが基本感情として存在するとしており，それぞれの感情に対しての強弱，また 2 つの基本感情を組み合わせることで生まれる 2 次感情，3 次感情があるとしている．このように，感情は定義がさまざまである．また，本論文でテーマにしている笑では，表現される感情と表現されない感情があると考えた．そのため，エクマンやプルチックの感情の定義や心理学の図書 [12]，[13] 参考にしつつ自らが収集したツイートと照らし合わせながら定めた．また，人がどのような時に笑うのかと関わりがあると考え [14]，[15] も参考にした．定めた感情ラベルは，大枠をポジティブ，ニュートラル，ネガティブ，ネガティブ＋ポジティブとし，ネガティブは攻撃的，非攻撃的，自虐に細分化した．

5.1　ポジティブに関して

エクマンの研究では幸せ，プルチックの感情輪では喜び，信頼がポジティブ感情だと捉えることができる．

ネガティブ感情に関しては，ある程度特定の状況を想定しやすい．お化け屋敷と聞けば恐れ・恐怖を連想する人は多いだろう．大事なものを壊されたと聞けば怒り，もしくは悲しみを連想するだろう．しかし，幸せなどは，特定の状況を想定しにくい．友達とゲームをしている時，恋人とデートをしている時，おいしいものを食べている時，いずれの場合でも幸せであると感じることができるだろう．しかし，これらが同一の感情を持っているとは考えにくい．

ポジティブ感情を細分化するためには，その状態・状況を細かく見る必要があり，困難である．そのため，本研究では細分化は行わず，ポジティブであると定義した．

5.2　ネガティブに関して

ネガティブ感情は，エクマンの定義でもプルチックの定義でも，悲しみ，恐れ，嫌悪，怒りの 4 つが定義されている．一方で，ポジティブ感情は，エクマンは幸せ，プルチックは喜び，信頼とどちらもネガティブ感情の方が多いことがわかる．しかし，悲しみや恐れなどの感情と雑魚すぎ笑という他人に対して攻撃的な感情を同じ感情ラベルとして扱うことが適切ではないと考えた．また，自虐的なテキストというのは少なからず他人を傷つけるための感情が表れているとは考えにくい．しかし，自身に対して攻撃的な文章だと捉えることができるため，自虐も感情ラベルとして分けることにした．

5.2.1　攻撃的に関して

本論文において，攻撃的とは，[16] を参考に，発言者以外の誰か・何か・事象に対して，軽蔑や何かを傷つけたい・破壊したいというような感情と定義している [16]．

ここでもう一つ，悪意について話をする．攻撃的であるとしても，必ずしも陥れよう，傷つけようとしているわけではないことを留めておいてほしいためである．ただし，本論文においては，悪意の有無に関係なく攻撃的であると定義している．例えば，怒りという感情は，一口に悪意のある感情といえるだろうか．多くの人は，怒りという言葉を聞くと，感情に任せて，ものに当たったり，人に暴力を振るったりという場面を想像する人は多いだろう．しかし，大切なものを傷つけられて，傷つけた人に対して，怒りを覚えるというのは悪意のある感情だと感じる人は少ないだろう．このように，負の感情であっても，一概に悪意があるとは言えないことがわかる．しかし，悪意があるかどうかの判断は難しいと考え本論文では考慮しないことにした．

具合的な例を挙げる．たかがアニメでよくそこまで熱くなれるな (笑) というテキストにおいては，アニメで熱くなれる人を軽蔑していることがわかる．ブログの内容，おもっきし下品で不愉快でクソワロタ．面白いと思って？高飛車に書いてるのだろうけど，スベっとるぞwwwフツーに胸糞w一般人の愚痴だったらいいけど，政治家なん？？？笑，というテキストにおいては，ブログを書いた人に軽蔑するような発言をしている．結局-300 だよ笑まじで考えて動けよな，特にデュオのや

つら，というテキストにおいては，デュオの人たちに怒りのような攻撃的な表現をしている．

5.2.2　非攻撃的に関して

これは恐れや悲しみなどの上記の攻撃的ではなくネガティブな感情であるもの，かつ後述の自虐的なものではないものと定義している．例えば，とりあえず今週の授業あと土曜日の受ければ終了な所まで終わらせた疲れたわ火曜 5 は教科書届かない限りはどうしようもない笑，というテキストにおいては，発言者が疲れている事に加え，教科書が届かないことが読み取れる．これは発言者にとって不都合な出来事だということがわかる．

5.2.3　自虐に関して

これは自分に対しての攻撃的な感情だと考え定義した．もの無くした笑，というテキストにおいては，悲しかったこと，つまり，非攻撃的な感情が表れている．しかし，もの無くした自分バカすぎ笑，というテキストにおいては，わざわざ自らをバカだと下げるような感情を表現している．

第 6 章　評価実験

データセットの作成はクラウドソーシングを用いて行った．過半数の判定者が同じ判定をしたものをラベルとして採用した結果，ニュートラルが 11772，攻撃的が 958，非攻撃的が 1686，ポジティブが 8453，自虐が 850，ポジティブ＋ネガティブが 906 の合計 24,625 のデータを収集することができた．この収集結果を受けて，私の感覚とは少し異なる結果となった．ネガティブに含まれる攻撃，非攻撃的，自虐のデータ，特に攻撃的，自虐のデータがもう少し多くなると考えていた．普段，私も Twitter を利用しているが，攻撃的なツイートや自虐的なツイートをしている人が多いという印象を持っているためである．Twitter が何でもないようなことを呟く場であること，日々起こったことや面白かったことを呟いている人が多いということから，特に感情を感じないニュートラルや楽しい，幸せのような感情を持つポジティブなツイートが多いという結果になったのだと私は考える．

作成したデータセットを用いて実験を行った．教師データを 80% とテストデータを 20% に分けている．本論文では，笑を含む文章がどのような感情を持っているかを機械学習を利用してどの程度判定可能かをを評価値を用いて評価を行う．ポジティブ，ネガティブ，ニュートラルの大枠 3 つでの分類とニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ポジティブ＋ネガティブの 6 つの分類を行う．また，データに偏りがあるため，少ないデータに関してはデータの水増しを行い，結果にどのような影響をもたらすかの実験も行った．

6.1　実験 1：大枠 3 つでの分類

本節では，ニュートラル，ネガティブ，ポジティブの大枠 3 つでの分類とデータの偏りを水増しによって行った場合の評価値の変化を確認する．

6.1.1　実験内容

収集したデータのうち，攻撃的，非攻撃的，自虐の 3 つのラベルをネガティブなラベルデータとして扱う．各ラベルのデータ数は表 6.1 に示す．ポジティブ＋ネガ

ティブは全体のデータに対して大枠として捉えるには非常に少ないと考え，この実験では考慮しないことにした．

6.1.2 結果・考察

データを水増しする前の実験結果を表 6.2 に，データを水増しした後の実験結果を表 6.3 に示す．Precision， Recall，F 値については，ニュートラル，ネガティブ，ポジティブのそれぞれを正とした時の数値を示している．

まず，水増し前の Accuracy は 0.690，水増し後の Accuracy は 0.705 となっており，どちらもランダムな予測をした場合の 33% よりは，高い値を取っている．そのため，水増し前の結果において，ネガティブが予測されないために，すべての評価値が 0.0 になっている．これはデータの偏りを考慮しておらず，学習する際にニュートラルかポジティブのどちらかを予測すれば良いということを学習したと考えられる．その点，水増しを行った場合，ネガティブ自体が予測される可能性があることがわかる．水増しを行うことは，Accuracy を大きく向上させることはない

表 6.1　大枠時のデータ数

	ニュートラル	ネガティブ	ポジティブ	総データ
水増し前	11772	3494	8453	23719
水増し後	11772	11772	11772	35316

表 6.2　実験 1: 大枠時のデータでの評価値（水増し前）

	評価値	ニュートラル	ネガティブ	ポジティブ
水増し前	Accuracy	0.690		
	Precision	0.670	0.0	0.726
	Recall	0.862	0.0	0.722
	F 値	0.754	0.0	0.724

ものの，有効であると考える．

また，ニュートラルの Recall の値に注目すると，0.862 から 0.751 と小さくなってしまっていることがわかる．逆に，Precision の値は 0.670 から 0.745 と大きくなっている．これはネガティブな感情を持つテキストを学習したことで，ニュートラルというラベルが付与されているものの，過半数の投票でラベルを決定しているために，微弱なネガティブ感情を持つテキストや，人によってはネガティブなテキストであると判断するような，評価者間で意見が分かれるようなテキストをネガティブであると判定したために，Recall は小さな値となり，Precision は大きな値になったと考える．ポジティブの値がほとんど変化しないのは，ポジティブとネガティブがはっきりと区別できる感情であり，予測する際に大きな影響がないためだと考える．

表 6.3　実験 1: 大枠時のデータでの評価値（水増し後）

	評価値	ニュートラル	ネガティブ	ポジティブ
水増し後	Accuracy	0.705		
	Precision	0.745	0.493	0.723
	Recall	0.751	0.448	0.741
	F 値	0.739	0.549	0.707

6.2　実験 2: 細分化した 6 ラベルでの分類

本節では，ニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ポジティブ＋ネガティブの私が定義した 6 ラベルでの分類とデータの偏りを水増しによって行った場合の変化を確認する．

6.2.1　実験内容

6.1 節とは異なり，攻撃的，非攻撃的，自虐，ポジティブ＋ネガティブはそれぞれ 1 つのラベルとして扱う．各ラベルのデータ数は表 6.4 に示す．

6.2.2　結果・考察

データを水増しする前の実験結果を表 6.5 に，データを水増しした後の実験結果を表 6.6 に示す．Precision， Recall，F 値については，ニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ポジティブ＋ネガティブのそれぞれを正とした時の数値を示している．

6.1 節と同様に水増し前の結果はデータの偏りのためにニュートラルかポジティブのみ予測する結果となった．水増し後において，ニュートラル，ポジティブ以外のラベルも予測はされるものの 6.1 節と比べると評価値はいずれも小さい．あまりにもデータの偏りが大きいと，水増しを行うだけではデータの偏りの問題は解消できないことがわかる．非攻撃的，自虐などのデータ数が少ないものに関しては，集中的にデータを集める必要があると考える．もしくは，定義した 6 ラベルの分類を一度に行うのではなく，6.1 節で作成した分類機と攻撃的，非攻撃的，自虐のみを

表 6.4　6 ラベル時のデータ数

	ニュ [1]	攻撃	非攻撃	ポジティブ	自虐	ポジ＋ネガ [2]	総データ
水増し前	11772	958	1686	8453	850	906	24625
水増し後	11772	11772	11772	11772	11772	11772	70632

[1] ニュートラル

[2] ポジティブ＋ネガティブ

分類する分類機を作成し，2 段階に分けて分類を行う必要があると考える．しかし，攻撃の評価値に着目すると，元々のデータ数が少ない中ではどの評価値においても高い値を取っていることがわかる．他のラベルに比べると，私が定めた感情の中でも特に他の対象に向けられやすく，表現や用いられる語もかなり特徴的な感情であると考えている．5.2.1 節でも挙げている例文のように，テキストを見ただけでもかなり対象に対して厳しい言い方をしていることがわかる．また，この例文に限らず，不適切な言葉を用いているものも存在する．そのため，データの種類自体は少ないもののかなりわかりやすい特徴を持っているために，元々のデータ数が少ない中では高い評価値が出ていると考える．

表 6.5　実験 2:6 ラベル時の評価値（水増し前）

	評価値	ニュ [1]	攻撃	非攻撃	ポジティブ	自虐	ポジ＋ネガ [2]
水増し前	Accuracy	0.667					
	Precision	0.648	0.0	0.0	0.707	0.0	0.0
	Recall	0.896	0.0	0.0	0.679	0.0	0.0
	F 値	0.752	0.0	0.0	0.694	0.0	0.0

[1] ニュートラル
[2] ポジティブ＋ネガティブ

表 6.6　実験 2:6 ラベル時の評価値（水増し後）

	評価値	ニュ [1]	攻撃	非攻撃	ポジティブ	自虐	ポジ＋ネガ [2]
水増し後	Accuracy	0.619					
	Precision	0.728	0.423	0.222	0.652	0.235	0.133
	Recall	0.862	0.456	0.218	0.680	0.239	0.128
	F 値	0.716	0.439	0.220	0.665	0.235	0.133

[1] ニュートラル
[2] ポジティブ＋ネガティブ

第 7 章　おわりに

本論文では「笑」という 1 つの感情語が持つ多様な感情表現の推定を目的とし，機械学習を用いることを提案した．同じ感情語であるにも関わらず，友達と久しぶりに会った笑のようなポジティブ，たかがアニメでよくそこまで熱くなれるな (笑) のような攻撃的な感情などの多様な感情表現に用いられている．そのため，感情を表現・強調するものとしては曖昧であることが問題である．そこで BERT を用いた機械学習により，どのような感情をもつかの判定を行う．データセット作成には，Twitter のツイートを用いた．また，笑の付与されているテキストが持つ感情をニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ポジティブ＋ネガティブの 6 ラベルを定義し，クラウドソーシングにより作業者に指示を行い感情ラベルを付与することでデータセットを作成した．この作成したデータセットを用いて，BERT による分類機を作成し，評価実験を行った．評価実験はニュートラル，ネガティブ，ポジティブの大枠 3 ラベルと，ニュートラル，攻撃的，非攻撃的，ポジティブ，自虐，ポジティブ＋ネガティブの定義した 6 ラベルの場合を行い，データの偏りを考慮するために，少ないデータに関しては同じデータを増やし，水増しした場合の実験も行った．

評価実験の結果を記す．大枠時の 3 ラベルでの分類，定義した 6 ラベル時の分類，どちらの場合も Accuracy のみに注目すれば，ランダムな予測よりも良い値であることがわかった．また，データの水増しを行うことである程度，評価値に影響をもたらすことがわかった．一方で，データの偏りが大きくなりすぎると，特徴を捉えきれないため水増しを行っても評価値が大きく改善されるわけではないという問題点がある．非攻撃的，自虐などの少ないデータのみを収集する，ネガティブのみを学習する分類機を作成し 2 段階で分類を行うなどをすることで，同じ感情語を持つテキストの分類がより改善すると考える．

謝辞

本研究にあたって, 指導教員である鈴木准教授にはたくさんのご助言, ご指導いただきました．また，同じ鈴木研究室の皆様には本研究について参考となる様々なご意見をいただきました．産休に入られた秘書の井尾さんと佐野さんにはクラウドソーシングの手続きをするにあたって大変お世話になりました．高校からの友人たちには，苦しい時に相談に乗っていただいたり，たくさんの励ましをしていただきました．家族には経済的・心身的に支援してくださったり，夜遅くでもたくさんの相談や世間話に付き合っていただいたりしていただき，深く感謝を申し上げます．

参考文献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. 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)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[2] 山本千尋, 別所克人, 内山俊郎, 内山匡. 絵文字の語義抽出と役割の曖昧性解消. 人工知能学会研究会資料 知識ベースシステム研究会 89 回, p. 07. 一般社団法人 人工知能学会, 2010.

[3] 黒崎優太, 高木友博. Word2vec を用いた顔文字の感情分類. 言語処理学会第 21 回年次大会 (NLP2015), B3-3, 京都大学吉田キャンパス, 2015.

[4] 源田翼, 横井健, 山本和英ほか. 構成要素に着目した顔文字の意味分析. 第 78 回全国大会講演論文集, Vol. 2016, No. 1, pp. 563–564, 2016.

[5] 大町凌弥, 瀧下祥, 奥村紀之. 文章と顔文字の組み合わせによる感情推定. 人工知能学会全国大会論文集 第 31 回 (2017), pp. 2O2OS22a2–2O2OS22a2. 一般社団法人 人工知能学会, 2017.

[6] 吉田綾奈, 邱起仁, 櫨山淳雄ほか. 顔文字推薦のための感情を付与した顔文字データベースの構築. 研究報告グループウェアとネットワークサービス (GN), Vol. 2014, No. 35, pp. 1–6, 2014.

[7] 堀宮ありさ, 坂野遼平, 佐藤晴彦, 小山聡, 栗原正仁, 沼澤政信. Twitter における発話者へのリプライを用いたユーザ感情推定手法. In *DEIM Forum*, 2012.

[8] 若井祐樹, 熊本忠彦, 灘本明代ほか. 映画に対する実況ツイートの感情抽出手法の提案. 研究報告データベースシステム (DBS), Vol. 2013, No. 16, pp. 1–6, 2013.

[9] 松林圭, 五味京祐, 古川和祈, 松尾祐佳, 松原良和, 中村拓哉, 山下晃弘, 松林勝志ほか. Twitter 上に投稿された文章に基づく感情推定法とその応用に関する検討. 第 78 回全国大会講演論文集, Vol. 2016, No. 1, pp. 79–80, 2016.

[10] 中澤政孝, 亀井且有, 前田陽一郎ほか. Bert を用いた単文の感情極性推定手法の提案とその有効性. 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集第 36 回ファジィシステムシンポジウム, pp. 177–180. 日本知能情報ファジィ学会, 2020.

[11] 高津弘明, 安藤涼太, 松山洋一, 小林哲則. ニュース記事のタイトルと文の系列に対する感情分析. 人工知能学会全国大会論文集 第 35 回全国大会 (2021), pp. 2Yin508–2Yin508. 一般社団法人 人工知能学会, 2021.

[12] 杉浦義典（編）. 感情・人格心理学, 公認心理師の基礎と実践, 9 巻. 遠見書房, 初版, 5 月 2020 年. 野島一彦, 繁桝算男監修.

[13] 内山伊知郎監修. 感情心理学ハンドブック. 北大路書房, 第 1 版, 9 月 2019. 日本感情心理学会企画.

[14] 井上宏. 笑い学のすすめ. 世界思想社, 7 月 2004.

[15] 木村覚. 笑いの哲学. 講談社, 1 月 2021.

[16] 秦一士, 湯川進太郎編訳. 攻撃の心理学. 北大路書房, 初版, 4 月 2004. B. クラーエ原著,The Social Psychology of Aggression(East Sussex. UK:Psychology Press,2001) の翻訳書.

発表リスト

[1] 吉山隆聖，鈴木優『「笑」が持つ様々な意味の理解』東海関西データベースワークショップ, 2021.
[2] 吉山隆聖，鈴木優『「笑」が持つ様々な意味の感情推定と解釈の一致に向けた感情ごとの置換の自動化』第 14 回データ工学と情報マネジメントに関するフォーラム, 2022.