

BERT 及び絵文字を利用した日本語文における皮肉の検出

諏訪 光輔[†] 張 建偉^{††}

[†] 岩手大学総合科学研究科 〒020-8551 岩手県盛岡市上田 4 丁目 3-5

^{††} 岩手大学理工学部 〒020-8551 岩手県盛岡市上田 4 丁目 3-5

E-mail: [†]{g0319098,zhang}@iwate-u.ac.jp

あらまし 近年 SNS 等の普及により重要となっている評判分析において、判定精度の低下の原因となる皮肉文の検出手法について提案する。ここでいう皮肉は、肯定的表現を用いながらも実際は否定的感情を示す文、またはその逆であり、これらは文の極性判定精度を低下させる要因となる。現在多くの皮肉検出研究が行われているが、文法が複雑で文脈理解が難しい日本語を対象とした研究は少ない。また、テキスト以外の新たな表現方法が生まれ、単語ベースでの文脈理解には限界があると考えられる。そこで本研究では、Bidirectional Encoder Representations from Transformers(BERT)を用いた文脈理解、及び文中の絵文字を考慮した皮肉文の検出方法を提案する。日本語 Twitter データセットを用い、テキスト及び絵文字それぞれの特徴ベクトルを抽出する。これらを用いて皮肉文検出を行った結果、ベースラインと比べ優位な結果を得た。

キーワード 皮肉検出, 深層学習, Twitter, BERT, 絵文字

1 はじめに

近年、インターネットの急速な普及に伴い Twitter¹に代表される SNS や掲示板などで個人によるテキストベースの情報発信が可能となり、また誰もがそれを利用できるようになった。図 1 は、日本における近年の SNS の利用者数の推移を示している [1]。図 1 より、2020 年度には SNS 利用者数が約 8,000 万人と、かなりの人数が SNS を利用していることがわかる。

その反面、一部の悪意のあるユーザーによって特定の対象に対する誹謗中傷、暴言などが増え、それら全てを漏れなく取り締まることは現状ほぼ困難である。これより、機械学習等を用いて文の意図や感情を自動で分析し活用する技術の開発が求められている。しかし、このように文のネガポジ分類を行うにあたって、「皮肉」を機械が理解することは簡単ではない。皮肉は文中で使われている単語の意味と実際に著者が示す意図に相違があることにより、既存の技術でも判定が難しく、AI に皮肉

を理解させることは自然言語処理における重要な課題となっている。

これより、多くの言語を対象とし、テキスト極性判定問題における皮肉文検出手法が盛んに研究されている。しかし、現在ほとんどの皮肉文検出研究は英語テキストを対象としており、日本語を対象とした研究は少ない。また、日本語を対象とした研究はテキスト以外の表現方法を考慮したものは筆者の知る限りでは存在しない。そこで、本研究では 2019 年に Jacob Devlin らによって発表された、BERT と呼ばれる文章の文脈を読むことが可能なモデルをベースとし、さらに文中に存在する絵文字から特徴ベクトルを取得することでテキスト極性判定タスクに影響を及ぼす皮肉を判別することを目的とした皮肉検出手法を提案する。

本論文の構成は以下の通りである。第 2 節では既存の皮肉文検出の研究について紹介する。第 3 節では本論文における皮肉の定義、および本論文において提案する皮肉文検出手法について説明し、第 4 節では評価実験とその結果について述べる。第 5 節では実験結果についての考察を述べ、第 6 節にて本稿をまとめる。

2 関連研究

皮肉文検出は文の極性判定問題の精度向上に役立つとされ、様々な手法 [2] が試みられている。Poria [3] らは、Convolutional Neural Network(CNN) を用いた皮肉文検出手法を提案し、皮肉文検出に有効であることを確認している。また、Zhang [4] らは、深層学習モデルをベースとし文脈を考慮した皮肉文検出手法を提案し、文脈理解による皮肉文検出の精度向上を達成している。

また、近年では Transformer [5] や BERT などの新たな言

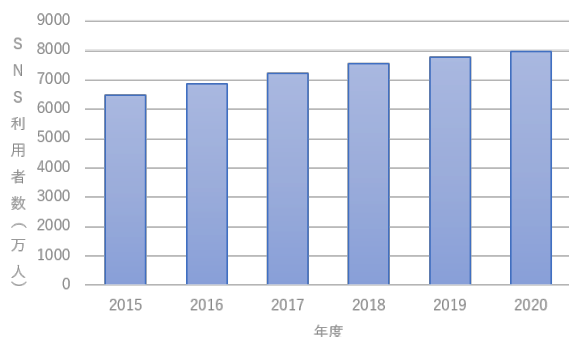


図 1 SNS の利用者数推移

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

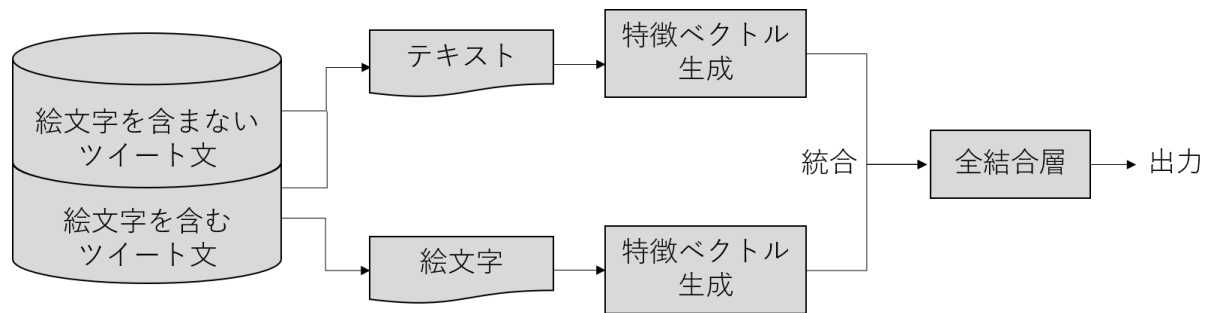


図2 提案手法の流れ

語モデルを用いた皮肉文検出手法の研究も行われている。Potamias [6] らは、Transformer と呼ばれるニューラル機械翻訳モデルを用いた皮肉文検出手法を提案している。また Zhang [7] らは、BERT を用いアラビア語を対象とした皮肉文検出手法について提案した。さらに、Zhang [8] らは、BERT 及び XLNet [9] モデルを用いた皮肉文検出手法を提案している。

一方、テキスト以外の表現も考慮した皮肉文検出手法の研究も存在する。Schifanella [10] らは、複数のソーシャルメディアにおいてテキストと画像の特徴を考慮し皮肉を検出する手法を提案している。また、Cai [11] らは、Twitter 上の投稿に対しテキスト、画像、画像属性の3つから特徴を得た後、皮肉の投稿を識別する手法を提案し、皮肉文検出への有用性を確認している。さらに Chaudhary [12] らの研究によって、英語テキストにおいて絵文字の特徴が皮肉文の検出に有効であることが示されている。

ここまで紹介した研究はいずれも英語のテキストを対象としている。一方、日本語を対象とした皮肉文検出手法は少ない。肥合 [13] らは、Twitter 上の日本語投稿を対象とし、LSTM 及び Attention に加えて関係ベクトルと呼ばれる特徴量を加え、ベースラインを上回る結果を出している。また、魚住 [14] らは、皮肉検出において感情生起要因を特徴量として用いることの有効性を検証している。しかし、これらの研究はテキストのみを対象としており、テキスト以外の表現方法が考慮されていない。本研究は BERT を用い文脈を理解することで日本語における皮肉文検出の精度を向上させ、さらに日本語皮肉文検出においてテキスト以外の表現を考慮することの有効性を検証することを目的とする。

3 皮肉の定義と提案手法

3.1 皮肉の定義

肥合 [15] らの研究では、皮肉の定義を「否定的な意味を伝える肯定表現」としている。この定義を参考にし、本研究においては表1で示す2つを皮肉と定義する。

表1 本研究における皮肉の定義

1. 字面だけみると一見肯定的（または中立）な文に見えるが、実際はある対象に対する否定的な感情を示している文
2. 字面だけみると一見否定的（または中立）な文に見えるが、実際はある対象に対する肯定的な感情を示している文

表1の定義をもとに、Twitter 上のデータを収集し、実験を行う。しかし、実際のデータセットにおいてはほとんどが定義1に該当するものであった。定義1に該当する皮肉文の例を以下に示す。

・「某ショッピングモールのフードコートにっていますが、民度の低さが肌に合っていて心地よいです」

この例では、「肌に合っていて心地よい」という肯定的な表現を用いて、「民度の低さ」に対する否定的な感情を示している皮肉文であることが読み取れる。肯定的表現と否定的文言が混在しており、文の極性判定の精度に影響を与えと考えられる。

・「毎年毎年ハズレを連れてくるスカウトを表彰したい」

この例では、「毎年毎年ハズレを連れてくるスカウト」に対しての否定的感情が読み取れるが、文末に「表彰したい」と肯定的な表現が用いられている皮肉文であることが読み取れる。この文も極性判定の精度に影響を与えと考えられる。

以上の定義に基づいて、データセットの構築を行い、皮肉文検出手法の評価実験に用いる。

3.2 提案手法

提案手法の流れは図2の通りである。提案手法の概要を以下に述べる。まず、分析対象となるデータに含まれている絵文字を抽出し、テキストと絵文字に分割する。ここでいう絵文字は、Unicode 13.1²によって定義されている絵文字のうち、本手法で用いる学習済みモデルによって分散表現を得られる1,661種の絵文字である。

次に、分割したテキストと絵文字それぞれについて特徴ベクトルを取得する処理を行う。テキストの特徴ベクトルを取得する方法として、BERT を適用する。BERT は、Devlin [16] ら

2: <https://unicode.org/>

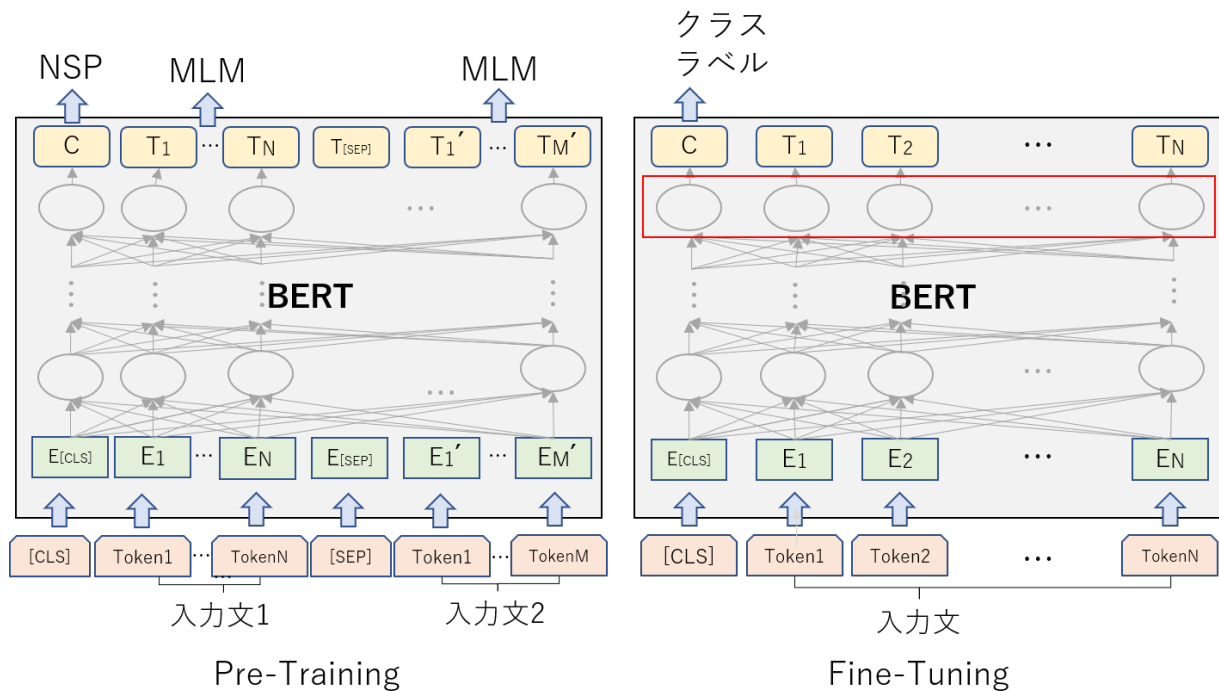


図3 BERTの構造

により開発された自然言語処理モデルであり、翻訳、文書分類等様々なタスクにおいて既存手法に比べ大きく優位な結果を残した「文脈を理解できる」モデルである。詳細は3.3節にて述べる。また、絵文字の特徴ベクトルを取得する方法として、word2vecを適用する。これについては3.4節で述べる。

最後に、取得したテキスト及び絵文字の特徴ベクトルを連結し、全結合ニューラルネットワークを用いて教師あり学習を行う。

3.3 Bidirectional Encoder Representations from Transformers(BERT)

BERTは、Google社³の研究者であるDevlin[16]らによって2019年に発表された自然言語処理モデルであり、「文脈を理解する」ことが可能な言語モデルである。BERTはその汎用性の高さから様々なタスクへの転移学習が可能であり、翻訳、文書分類、質問応答等様々な自然言語処理タスクにおいて、当時最高スコアを更新した。

BERTモデルの概要について説明する。BERTの構造については、図3の通りである。BERTの学習には、事前学習及びファインチューニングの2段階がある。事前学習時は、1つの文または2つの文をラベルなしデータとして入力し、汎用的な言語モデルを獲得する。ファインチューニングでは、事前学習で得た重みを各ノードの初期値とし、ラベル付きのデータを用いて教師あり学習を行いタスクに対応させる。

事前学習は、Masked Language Modeling(MLM)及びNext Sentence Prediction(NSP)の2手法により行われる。MLMでは、入力されたトークンの15%をMaskトークンに変換し学習

することで、元々何というトークンであったかを予測する穴埋め問題を解く。NSPでは、2文を選択しそれらが連続した文であるかを予測する。

ファインチューニングでは、ラベル付きデータを用いて従来の深層学習モデルと同様にタスクを学習させ重みの更新を行う。

また、BERTはタスクへのファインチューニングだけでなく、中間層の重みを抽出することでテキストの特徴ベクトルとして用いることが可能である。図3の赤枠で囲まれた部分は、BERTの最終層(出力層)の1つ手前の層を示しており、この層の出力値を入力文の特徴ベクトルとして用いることが可能である。

本稿における提案手法では、BERTによって得られた特徴ベクトル(1,024次元)を用いて学習を行う。

3.4 word2vec

word2vecは、Google社の研究者であるMikolov[17]らによって2013年に発表された自然言語処理モデルであり、それまで文字や単語をベクトルに変換する手法として用いられてきたone-hotとは異なり、単語の分散表現(Word Embedding)を獲得できるモデルである。one-hotベクトルは、ベクトルの要素のうち1つが1であり、他がすべて0であるベクトルのことである。この表現方法では、ベクトルの次元数が単語の総数になってしまうため膨大な計算量となる問題があった。対して分散表現は単語をベクトル空間に埋め込み、ベクトル空間上の点として表現することで次元数を抑え、かつ単語同士の関係まで表現することができる。有名な例として、「王様」を表すベクトルから「男性」を表すベクトルを減算し、「女性」を表すベクトルを加算することで「女王」を表すベクトルとなる。このように、単語間の関係性を加味したベクトル空間への埋め込みが

3: <https://about.google/>

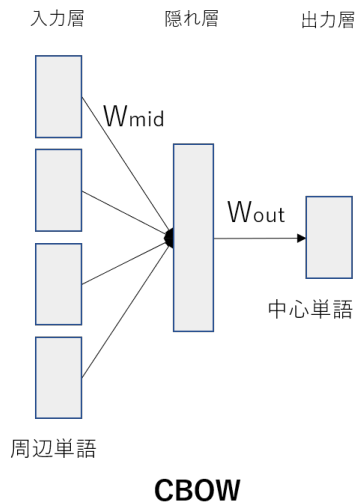


図 4 CBOW の構造

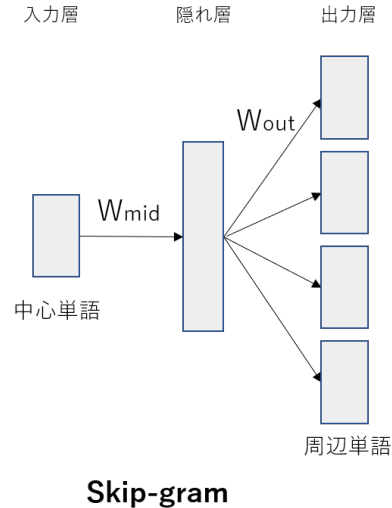


図 5 Skip-gram の構造

行われる。

word2vec の構造に関して説明する。word2vec は、図 4 または図 5 に示すニューラルネットワークモデルのうちいずれかで構成される、それぞれ CBOW(Continuous Bag-of-Words) 及び Skip-gram と呼ばれる手法である。

CBOW はコンテキストからターゲットを推測することを目的としたニューラルネットワークモデルである。CBOW では、ある 1 文のうち 1 単語の周辺単語それぞれの one-hot ベクトルを入力とし、中心単語の one-hot ベクトルが出力となるモデルを学習する。この学習により、CBOW は周辺単語から中心単語を予測できるようになり、隠れ層の重み W_{mid} は単語の分散表現となる。

Skip-gram は、CBOW とは逆にターゲットからコンテキストを推測することを目的としたニューラルネットワークモデルである。Skip-gram では、CBOW とは反対に中心単語の one-hot ベクトルを入力とし、その周辺単語それぞれの one-hot ベクトルを出力とするモデルを学習する。この学習により、Skip-gram はある単語からその周辺単語を予測できるようになり、隠れ層の重み W_{mid} は単語の分散表現となる。

提案手法では文中に含まれる絵文字に対して word2vec を用いて分散表現を獲得し、それを絵文字の特徴ベクトル (300 次元) として用いる。2 つ以上の絵文字が含まれている場合は、各ベクトルの平均値をとることによってその絵文字の特徴ベクトルとする。

4 評価実験

4.1 データセットと前処理

評価実験に用いるデータセットについて説明する。対象とするデータは、Twitter 上の日本語投稿文のうち、単体で皮肉か否かを評価できるものとし、Twitter API⁴を用い特定のクエ

リを指定しデータの収集を行う。収集期間は 2010 年 1 月から 2020 年 10 月である。多くの皮肉文検出の先行研究では、ツイート収集のクエリとして「#sarcasm」を用いている。本研究においても、「#皮肉」及び「(皮肉)」をクエリとして収集し、12,070 件のデータを得た。

次に、収集したデータに表 2 の通り前処理を施した。本研究では、そのツイート単体で評価可能なものを対象とするため、文脈以外の情報 (外部サイト、返信内容等) の情報が必要となるものは除いた。前処理の結果、データ数は 6,020 件となった。

そして、選出したデータに対しクラウドソーシングを利用しツイートが皮肉文か否かを判定してもらうことにより、最終的なラベリングを行った。判定基準は 3.1 節において述べた皮肉の定義とし、Yahoo!クラウドソーシング⁵を利用し 1 ツイートに対し 3 名のワーカーに判定してもらう。そして 2 名以上に皮肉であると判定されたツイートを真の皮肉文としてラベリングした。最終的に、皮肉ツイートデータは 3,025 件となった。

また、上記とは別に Twitter 上からランダムに収集した日本語投稿文に皮肉ツイートデータと同様の前処理を行い得られた、真データと同数の 3,025 件を非皮肉文データとして用いた。得られたデータセットの内訳は表 3 の通りである。また、表 4 に、データセットの例を示す。

表 2 前 処 理

テキスト内に@ (メンション) を含むツイートを削除
テキスト内に URL を含むツイートを削除
皮肉の意味をもつ単語 (皮肉, sarcasm, irony 等) を含むツイートを削除
「#皮肉」及び「(皮肉)」部分を削除

4 : <https://developer.twitter.com/>

5 : <https://crowdsourcing.yahoo.co.jp/>

表 3 データセットの内訳

	絵文字あり	絵文字なし	計
皮肉文	705 件	2,320 件	3,025 件
非皮肉文	705 件	2,320 件	3,025 件
計	1,410 件	4,640 件	6,050 件

表 4 データセットの一例

皮肉文	賞味期限 1 週間過ぎた牛乳を飲ませるなんてママ、 アンタ最高だね！
皮肉文	自分で見て感じたこと以外のことつまりニュースとかネットの ことを簡単に信用できる人って素直にすごいと思うよ
非皮肉文	モニターと LAN ケーブル買いませんか
非皮肉文	青春って感じの写真も映像も撮りたい… 海もいいし教室もいいし公園もいいし…。

表 5 評価実験リスト

実験	対象データ	手法
1-1	全データ (6,050 件)	ベースライン
1-2	全データ (6,050 件)	BERT ファインチューニング
2-1	絵文字を含むデータ (1,410 件)	提案手法 (絵文字を考慮しない)
2-2	絵文字を含むデータ (1,410 件)	提案手法 (絵文字を考慮する)

4.2 実験環境

実験環境には、Google Colaboratory⁶を用いた。Google Colaboratory は、Google 社が研究プロジェクトの 1 つとして公開している、ブラウザ上で Python プログラムを実行できるサービスである。Python の主要なライブラリは構築済みであり、GPU、TPU の使用が可能であるため、深層学習などの大規模な計算を行うことが容易になる。実験はメモリ 13GB RAM、Tesla K80 GPU、TPUv2 の環境を用いて行った。

また、BERT の事前学習済みモデルとして京都大学の黒橋研究室より公開されている BERT 日本語 Pretrained モデル⁷を用いた [18]。また、word2vec の絵文字学習済みモデルとして emoji2vec⁸を用いた [19]。

4.3 実験内容

評価実験について説明する。表 5 は実施する実験のリストである。実験 1(1-1, 1-2) の目的は、皮肉文検出における BERT の有効性の検証、そして実験 2(2-1, 2-2) の目的は皮肉文検出において絵文字を考慮することの効果を検証することである。実験 1 では全データ (6,050 件) を用い、実験 2 では絵文字の含まれたデータ (1,410 件) のみを用いる。

実験 1-1 ではベースラインとして、既存研究で多く用いられている Bi-LSTM 及び Attention を用いたモデルで実験を行う。バッチサイズは 16、エポック数は 10 とし、Bi-LSTM の中間

層のユニット数は 512 である。

実験 1-2 では、BERT によるタスクへのファインチューニング実験を行う。ドロップアウト率は 0.1、学習係数は 1.0×10^{-5} 、バッチサイズは 16、エポック数は 10 である。

実験 2-1 では提案手法を用いるが、word2vec を用いた絵文字の特徴ベクトル取得を行わず BERT によるテキスト特徴ベクトルの取得のみ行う。ドロップアウト率は 0.1、中間層はユニット数 1,024, 512, 256, 128 の 4 層、バッチサイズは 8、エポック数は 30 である。

実験 2-2 ではテキストと絵文字の両方を考慮した提案手法での実験を行う。パラメータは実験 2-1 と同様である。

また、実験は層化 5 分割交差検証 (stratified 5-fold cross-validation) により行った。評価指標には以下の数式で示される適合率 (precision)、再現率 (recall)、F 値 (F-measure) を用いた。真陽性を TP、偽陽性を FP、偽陰性を FN とする。

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

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

$$F - measure = \frac{2Recall \cdot Precision}{Recall + Precision}$$

4.4 実験結果

4.3 節で述べたように、まず日本語皮肉文検出における BERT の有効性を検証することを目的とした実験 1 を行った。実験 1-1 はベースライン、実験 1-2 は BERT によるファインチューニングである。結果を表 6 に示す。表 6 に示す通り、BERT を用いてファインチューニングを行った実験 1-2 の結果が、ベースラインを上回る結果となった。

そして、日本語皮肉文検出における絵文字を考慮することの有効性の検証を目的とした実験 2 を行った。実験 2-1 は提案手法であるがテキストのみを考慮し、実験 2-2 ではテキストに加えて絵文字も考慮した手法である。結果は表 7 の通りである。表 7 に示す通り、絵文字を考慮した実験 2-2 の結果が、実験 2-1 の結果を上回る結果となった。

また、各実験結果における混同行列を表 8、表 9、表 10、表 11 に示す。

表 6 実験 1 結果

	適合率	再現率	F 値
実験 1-1	0.78	0.79	0.79
実験 1-2	0.80	0.84	0.82

表 7 実験 2 結果

	適合率	再現率	F 値
実験 2-1	0.73	0.75	0.74
実験 2-2	0.79	0.80	0.80

6 : <https://colab.research.google.com/>

7 : <http://nlp.ist.i.kyoto-u.ac.jp/>

8 : <https://github.com/uclnlp/emoji2vec>

表 8 実験 1-1 の混同行列

		予測結果	
		皮肉	非皮肉
正解ラベル	皮肉	623	152
	非皮肉	226	549

表 9 実験 1-2 の混同行列

		予測結果	
		皮肉	非皮肉
正解ラベル	皮肉	648	127
	非皮肉	159	616

表 10 実験 2-1 の混同行列

		予測結果	
		皮肉	非皮肉
正解ラベル	皮肉	103	38
	非皮肉	35	106

表 11 実験 2-2 の混同行列

		予測結果	
		皮肉	非皮肉
正解ラベル	皮肉	111	30
	非皮肉	33	108

5 考 察

日本語皮肉文検出における BERT の有効性と、日本語皮肉文検出における絵文字を考慮することの有効性について考察する。

5.1 日本語皮肉文検出における BERT の有効性

表 6 より、BERT のファインチューニングを用いた皮肉文検出手法の適合率、再現率、F 値はそれぞれ 0.80, 0.84, 0.82 であり、これはベースラインの 0.78, 0.79, 0.79 を上回っていた。また、実際に BERT を用いることによってベースラインを用いた場合から判定結果が変化した件数を表 12 に示す。この表より、ベースラインでは誤って判定されていたものが BERT によって正しく判定された件数が、ベースラインで正しく判定されていたものが BERT を用いることで誤判定となった件数を上回っており、この結果より BERT が日本語文においても皮肉文検出に有効であると考察できる。

5.2 日本語皮肉文検出における絵文字を考慮することの有効性

表 7 より、絵文字を考慮した皮肉文検出手法の適合率、再現率、F 値はそれぞれ 0.79, 0.80, 0.80 であり、これは絵文字を考慮しない場合の 0.73, 0.75, 0.74 を上回っていた。また、実際に絵文字を考慮することによって絵文字を考慮しない場合から判定結果が変化した件数を表 13 に示す。この表より、絵文字を考慮しない場合では誤って判定されていたものが絵文字を

表 12 BERT を用いることで結果が変化した件数

実験 1-1		実験 1-2		件数	
予測結果	正誤	予測結果	正誤		
非皮肉	×	皮肉	○	88	216
皮肉	×	非皮肉	○	128	
非皮肉	○	皮肉	×	61	124
皮肉	○	非皮肉	×	63	

表 13 絵文字を用いることで結果が変化した件数

実験 2-1		実験 2-2		件数	
予測結果	正誤	予測結果	正誤		
非皮肉	×	皮肉	○	15	27
皮肉	×	非皮肉	○	12	
非皮肉	○	皮肉	×	10	17
皮肉	○	非皮肉	×	7	

考慮することによって正しく判定された件数が、絵文字を考慮しない場合に正しく判定されていたものが絵文字を考慮することで誤判定となった件数を上回っていた、この結果のみ考慮すれば絵文字が日本語文において皮肉文検出に有効であると考察される。しかし、表 13 から絵文字を考慮することによって予測結果が正しく変化した件数は 27 件であるのに対して誤った結果となった件数は 17 件あり、全てのデータにおいて絵文字が皮肉文の検出に効果的であったとは断定できない。原因としては、本研究において用いたデータセットのうち絵文字を含んでいるものが、表 3 で示す通り 1,410 件と少なく、学習データとしては少ないことであると考察される。

5.3 判定結果の例

実際に BERT により判定結果が変化した例を表 14、絵文字を考慮することで判定結果が変化した例を表 15 に示す。

絵文字を考慮することで誤判定へと変化した表 15 の 3, 4 文目に関して、前者は「マジックレベル 10 の豆アズ」に対する「カッコいい」、つまりプラスの感情を示す非皮肉文であることが読み取れるが、最後の泣いた顔の絵文字がマイナス感情を示しており誤って皮肉と判定されたのではないかと推測できる。また、後者はテキストのみでも十分皮肉文であることが理解できるが、絵文字を考慮することによって誤判定となった。プラスの感情を示す、拍手をする絵文字と腕で丸を示す人の絵文字によって文全体を通して肯定的であると判定されたと考察することもできる。また、絵文字を考慮することで誤判定となったデータに含まれる絵文字は、1 つではなく 2 つ以上のものが比較的多い傾向にあった。しかしデータセットの規模が小さいため、より大規模なデータセットでの検証が必要であると考え。

6 ま と め

本稿では、日本語テキストを対象とした新たな皮肉文検出手法の提案を行った。従来と比較してより文脈の理解に長けた BERT と呼ばれる言語モデルを日本語テキストに対して適用し、皮肉文検出精度の向上を図った。また、文中の重要な要素

表 14 BERT によって結果が変化した例

実験1-1		実験1-2		Tweet
予測結果	正誤	予測結果	正誤	
非皮肉	×	皮肉	○	某ショッピングモールのフードコートにっていますが、民度の低さが肌に合っていて心地よいです
皮肉	×	非皮肉	○	昔は卒アルに連絡先書いてありましたが今は書いてないのかもですね
非皮肉	○	皮肉	×	こんにちわ！日本語が下手ですからなにか間違いがあったら気軽に訂正してね！
皮肉	○	非皮肉	×	毎年毎年ハズレを連れてくるスカウトを表彰したい

表 15 絵文字を考慮することで結果が変化した例

実験2-1		実験2-2		Tweet
予測結果	正誤	予測結果	正誤	
非皮肉	×	皮肉	○	元検(犬🐶)が当選したんだってwww自民党の皆さんおめでとう！
皮肉	×	非皮肉	○	6月って無駄にお金使ってる気がするしめんどいことがよく起こる😓
非皮肉	○	皮肉	×	マジックレベル10の豆アズめちゃくちゃカッコいいから見てほしい😭😭🙏🙏
皮肉	○	非皮肉	×	他人を晒して机上の理論並べてる人自分を持ってて好きですよ👍👍♂

である絵文字を考慮し、その特徴をテキスト特徴と同時に用いることで検出精度の向上を図った。評価実験には Twitter 上の日本語投稿文を用い、クラウドソーシングによってより信頼性の高い皮肉文データセットを構築した。結果、どちらの手法においても従来よりも検出精度が向上する結果となった。

しかし、本研究では Twitter 上において「#皮肉」または「(皮肉)」を文中に含むツイートを収集したが、このようなツイートはそもそも少なく、その中でも絵文字を含むものは限られていた。皮肉か非皮肉かの正確なラベリング手法は、現状テキストを人間の目で1つ1つチェックする以外なく、大量のデータをすべてラベリングするのは困難を極める。したがって、より大規模かつ信頼性の高い皮肉テキストデータセットを構築するための効果的な手法を確立する必要があると考える。また、本研究における非皮肉データはランダムに収集したものをを用いており、大量の非皮肉データのラベリング方法を確立することも今後の課題であると考ええる。

また今後の研究要素として、文中における絵文字の重要度を考慮することにより、より精度の高い皮肉文検出ができるのではないかと考える。加えて、BERT 以降に発表され近年 BERT 以上に大きな成果を挙げている XLNet [9], RoBERTa [20], ALBERT [21] 等を適用することによって、より判定精度を高められるのではないかと考える。

謝 辞

本研究は JSPS 科研費 19K12230 の助成を受けたものである。

文 献

- [1] ICT 総研「2020 年度 SNS 利用動向に関する調査」, 2020. <https://ictr.co.jp/report/20200729.html>
- [2] Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman. Automatic Sarcasm Detection: A Survey. ACM Computing Surveys(CSUR), 50(5):73:1-73:22, 2017.
- [3] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, and Prateek Vij. A deeper look into sarcastic tweets using deep convolutional neural networks. In COLING, 1601-1612, 2016.
- [4] Meishan Zhang, Yue Zhang, and Guohong Fu. Tweet sarcasm detection using deep neural network. In COLING, pages 2449-2460, 2016.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, 5998-6008, 2017.
- [6] Rolandos Alexandros Potamias, Georgios Siolas, and Andreas - Georgios Stafylopatis. A Transformer-based approach to Irony and Sarcasm detection. In Neural Computing and Applications Volume 32, 17309-17320, 2020.
- [7] Chiyu Zhang, and Muhammad Abdul-Mageed. Multi-Task Bidirectional Transformer Representations for Irony Detection. In FIRE 2019, 391-400, 2019.
- [8] Cheng Zhang, Masashi Kudo, and Hayato Yamana. Evaluation of BERT and XLNet Models on Irony Detection in English Tweets. In DEIM Forum 2020 G1-4, 2020.
- [9] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V.Le. Xlnet: Generalized Autoregressive Pretraining for Language Understanding. In NeurIPS 2019, 5754-5764, 2019.
- [10] Rossano Schifanella, Paloma de Juan, Joel Tetreault, and Liangliang Cao. Detecting Sarcasm in Multimodal Social Platforms. In ACM on multimedia conference, pages 1136-1145, 2016.
- [11] Yitao Cai, Huiyu Cai, and Xiaojun Wan. Multimodal sar-

- casm detection in twitter with hierarchical fusion model. In ACL, pages 2506–2515, 2019.
- [12] Aditi Chaudhary, Shirley Anugrah Hayati, Naoki Otani, and Alan W Black. What A Sunny Day : Toward Emoji-Sensitive Irony Detection. In W-NUT, pages 212–216, 2019.
 - [13] 肥合智史, 嶋田和孝. 関係ベクトルを利用した皮肉の検出. 言語処理学会 第 24 回年次大会, pp.829–832, 2018.
 - [14] 魚住ゆい, 内田ゆず, 荒木健治. 皮肉検出における感情生起要因の有効性. 第 17 回情報技術フォーラム (FIT2018), pp.163-164, 2018.
 - [15] 肥合智史, 嶋田和孝. 評価表現に着目した皮肉文の分析と抽出. 火の国情報シンポジウム 2016, 4A-1, 2016.
 - [16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT, pages 4171-4186, 2019.
 - [17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. In ICLR Workshop, 2013.
 - [18] 柴田 知秀, 河原 大輔, 黒橋 禎夫. BERT による日本語構文解析の精度向上. 言語処理学会 第 25 年次大会, pp.205-208, 2019.
 - [19] Ben Eisner, Tim Rocktaschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. emoji2vec: Learning emoji representations from their description. In SocialNLP, 2016.
 - [20] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. CoRR abs/1907.11692. 2019.
 - [21] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. In 8th ICLR, 2020.