

# 多人数向けメッセージからの失礼表現の自動抽出

## Extraction of Impolite Expression from Messages to Multipersons

安藤律子 砂山渡\*

Ritsuko Ando Wataru Sunayama

広島市立大学大学院 情報科学研究科

Graduate School of Information Sciences, Hiroshima City University

**Abstract:** It is desired that messages don't become impolite when you send a message towards many unspecified persons. This paper aimed to extract the most likely sentence that contains impolite expression from message such as BBS and Twitter. The system uses the set of words that are more likely to be impolite expression to judge whether a message is impolite. The result of the system promotes the reconsideration of message contents to user. In addition, the result helps to avoid a contribution of the impolite expression.

### 1 はじめに

近年、文章を用いたインターネットコミュニケーションが、盛んに行われている。例えば、電子掲示板や Twitter が挙げられる。これらの利用者は、不特定多数の人に向け、様々な内容のメッセージを発信する。しかし、メッセージの発信者は、読みだけに徹しているの読者の存在に気がつきにくいことや、ネット特有の匿名性によって、無責任な発言や批判的な発言をしてしまう [1]。そういった状況下で、良好なコミュニケーションをとるためには、メッセージの発信者が、発信するメッセージの表現を読者に対し、失礼にならないように配慮をすることが望まれる。しかし、メッセージの表現が、知らないうちに読者への配慮が足りない表現となることや、発信するメッセージ内容が、大量な場合、失礼な表現が含まれる文の存在に気がつきにくく、見落とし易くなることが考えられる。そこで本研究では、電子掲示板や Twitter 等、多人数向けメッセージの集合を対象に、失礼表現が含まれている可能性が高い文(以下、失礼文と記述する)を抽出することを目的とする。システムの利用者に、発信予定のメッセージ集合に含まれる失礼文の存在を提示し、使用者の失礼文の見落としを防ぎ、メッセージ内容の見直しをアシストする。

### 2 関連研究

#### 2.1 電子掲示板, Twitter に関する研究

本節では、電子掲示板, Twitter に関する研究について述べる。電子掲示板を対象としたコミュニケーションの雰囲気に関する研究 [2, 3] がある。これらの研究は、電子掲示板に投稿されている発言から、電子掲示板の雰囲気を判定し、システムの利用者に提示を行うものであった。また、Twitter から利用者個人の特徴を抽出する研究 [4, 5] がある。これらの研究は、投稿されているツイートから、ツイートした人物の特徴を得て、システムの利用者に提示を行うものであった。4つの研究を挙げたが、電子掲示板や Twitter に関する研究は、既に投稿された文から、読者に対し、有益な情報を与える研究が多かった。本研究では、電子掲示板や Twitter に文を投稿する前に、文を投稿する人に対して、その文が失礼文かどうかの情報を与える。

#### 2.2 文章の抽出に関する研究

本節では、文の抽出に関する研究について述べる。各研究で、対象とする文の抽出を行う研究 [6, 7] がある。これらの研究では、手がかり語や語句間の関係性、係り受け関係にある文節に着目し、対象とする文を抽出していた。本研究では、語句間の関係や文節係り受け関係には着目せず、文中で使用されている単語に着目して、失礼文の抽出を行う。

\*連絡先: 〒 731-3194 広島県広島市安佐南区大塚東 3-4-1

## 2.3 表現の判定に関する研究

本節では、表現の判定に関する研究について述べる。入力に対し、誹謗中傷表現を判定する研究 [8]、不満表現を判定する研究 [9]、有害情報を判定する研究 [10] がある。これらの研究は、人手で収集した単語に、表現の度合いを示す値を付与したり、収集した単語とは逆の意味を持つ単語を再度収集、比較したりすることで表現の判定を行うものであった。本研究では、これらの研究と同じく、アンケートをとり、人手で単語収集を行った。しかし、表現の判定には、失礼表現の度合いを示す値の付与や、失礼表現とならない単語の収集の手間を省き、その単語が 1 文に含まれているか、含まれていないかで行っている。

## 3 失礼文自動抽出システム

### 3.1 失礼文の定義

本研究で抽出の対象としている失礼文の定義について述べる。一般に失礼とは、他人に接する際の心得をわきまえていないこと、また、礼儀に欠けることを指す。すなわち、人間関係や社会生活の秩序を維持するために守るべき行動が欠けていることが失礼に当たると考えられる。そこで本研究では、『他者や物に対して、必要以上に否定的な表現、または蔑視的な表現』を失礼表現と定め、このような失礼表現を含む文を失礼文として定義する。

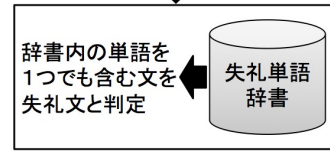
### 3.2 システム構成

本節では、システムの構成について述べる。図 1 にシステムの構成を示す。本システムは、入力された多人数向けメッセージの集合から、辞書内の単語を 1 つでも含む文を失礼文と判定し、色付けをして出力する。失礼文の判定には、失礼単語辞書を用いる。失礼単語辞書は、失礼表現となる可能性がある単語を集めた辞書である。3.4 項にて、失礼文を抽出する際に用いる失礼単語辞書の作成について述べる。

### 3.3 入力：多人数向けメッセージの集合

本節では、失礼文自動抽出システムへの入力について述べる。本システムには、入力として、多人数向けメッセージを記述したテキストを与える。本研究では、多人数向けメッセージを電子掲示板や Twitter 等、多くの人々に読まれる可能性がある場に投稿するメッセージとする。

入力：多人数向けメッセージの集合



出力：失礼文

図 1: 失礼文自動抽出システムの構成

### 3.4 失礼単語辞書の作成

本節では、失礼分の判定の用いる失礼単語辞書を作成した方法について述べる。以下、失礼単語辞書の作成した方法について、既存の辞書からの失礼単語の収集と失礼単語の追加の順に述べる。

#### 3.4.1 失礼単語の収集

失礼単語の収集について述べる。失礼単語は、失礼文の定義に沿った名詞、形容詞とし、単語感情極性対応表 [11] から収集する。単語感情極性対応表とは、ある単語が一般的に良い印象を持つか悪い印象を持つかを、感情極性値 -1 から 1 で表したものになる。この感情極性値が正の値ならば良い印象を、負の値ならば悪い印象を表している。一般的に悪い印象を持つ単語は、失礼表現となる単語の可能性が高いと考え、その単語を元にした。失礼単語は、単語感情極性対応表内の感情極性値が -0.48 以下の単語、14035 単語についてアンケートをとり、その結果から収集した。アンケートは、情報科学を専攻する大学生・大学院生の男女 6 人に、単語について、その単語を含む文が次の 4 つの項目のいずれに当てはまるかを評価してもらった。

- (1) ほぼ間違いなく失礼な文になる
- (2) 場合によっては失礼な文になる
- (3) 基本的に失礼な文にならない
- (4) 単語の意味がよく分からない

アンケートの回答に点数を割り振り、(1) の 1 回答につき 3 点、(2) の回答数が 1 つにつき 1 点とした。この時点で、1 単語当たりの回答人数や合計点から単語を振り分け、それぞれ以下の条件が当てはまる 10 種類の辞書 (辞書を構成する単語数) を用意した。

これらの辞書について、適切な辞書を決定するための予備実験を行った。10 種類の辞書を用いて失礼文を

表 1: 留意した辞書とアンケート結果との関係一覧

	(1) の人数	(2) の人数	単語数
辞書 1	3 人以上		325
辞書 2	3 人以上	または, 3 人以上	878
辞書 3	1 人以上		1068
辞書 4	1 人以上	または, 3 人以上	1380
辞書 5	1 人以上	または, 1 人以上	2014
辞書 6	2 人以上	または, 3 人以上	1052
辞書 7	1 人以上	または, 2 人以上	1644
辞書 8	人数×3 点	×1 点: 計 8 点以上	406
辞書 9	人数×3 点	×1 点: 計 5 点以上	744
辞書 10	人数×3 点	×1 点: 計 2 点以上	1115

表 2: 辞書の適合率, 再現率, 抽出正解文数

	適合率	再現率	抽出正解文数
辞書 1	0.71	0.38	33
辞書 2	0.56	0.48	42
辞書 3	0.56	0.48	42
辞書 4	0.52	0.51	45
辞書 5	0.45	0.51	45
辞書 6	0.57	0.51	45
辞書 7	0.49	0.52	46
辞書 8	0.62	0.41	36
辞書 9	0.50	0.41	36
辞書 10	0.61	0.53	47

抽出し, その適合率, 再現率を求めた. 抽出に用いた文は, 電子掲示板からの 200 文と Twitter から 200 文の計 400 文とした. 予備実験に使用した文は, 失礼文の数が極端にならないよう, 選択, 収集をした. このとき, 大学生 3 人に抽出に用いた文を読んでもらい, うち 2 人が失礼な文と判断した 88 文を正解文とした.

表 2 に, 作成した 10 個の辞書の適合率, 再現率, 抽出正解文数を示す. 表 2 から, 抽出正解文数が多く, 再現率が高い辞書は辞書 7 と辞書 10 であることがわかる. しかし, 適合率をみると辞書 10 の方が優れている. よって, 辞書 10 の単語を失礼単語辞書に用いることとした. 表 3 に, この時点で収集できた失礼単語, 名詞 942 単語と形容詞 173 単語, 計 1115 単語からそれぞれ一部抜粋したものを示す.

### 3.4.2 失礼単語の追加

前節の辞書では再現率の値が十分でなく, その理由は, 辞書の構築に用いた単語感情極性対応表がブログ記事を元に作られたものであり, 電子掲示板やつぶやきによく含まれる単語でも, 辞書内には含まれていない単語が見られたことによる. そこで, 電子掲示板と Twitter のメッセージから, 失礼表現を以下の手順によって追加した.

手順 1 電子掲示板と Twitter からメッセージの集合(合計 1000 文<sup>1</sup>)を収集し, 辞書 10 を用いて失礼文を抽出.

手順 2 失礼文と判定されなかった文を大学生 3 人に読んでもらい, その中から失礼な文を選択してもらう.

こうして収集した 1000 文から, 大学生 3 人に失礼単語となる可能性が最も高い単語を, 1 文につき, 1 単語

<sup>1</sup>失礼単語の収集が目的のため, 失礼文が多く存在しそうな文を収集した.

表 3: 収集した失礼単語 (一部抜粋)

品詞	失礼単語
名詞	生意気 最低 邪魔 不細工 低能 音痴 非常識 無神経 外道 非道 ろくでなし 厄介 不評 ごみ 下手
形容詞	つまらない みっともない あざとい 鈍い 陰気臭い とろい 薄気味悪い どぎつい 騒がしい

取り出してもらった. 結果として, 各文において 3 人中 2 人以上が抽出した単語が存在したため, 合計 1000 単語を追加の候補とした. その後, 収集した 1000 文の中で 2%以上の文から回答された 37 単語を失礼単語として, 辞書に追加した. 表 4 に追加した単語の一部を示す.

最終的に作成された失礼単語辞書は, 名詞 973 単語, 形容詞 179 単語の合計 1152 単語から構成される.

### 3.5 失礼単語辞書による失礼文の判定

前節で作成した失礼単語辞書を用いて, 入力として与えられた文が失礼単語辞書内の単語を 1 単語以上含んでいる文を失礼文として判定する.

### 3.6 出力: 失礼文の表示

システムの出力例を図 2 に示す. システムは, 失礼文と判定した文を橙色で色づけて出力する. また, 桃色で色づけされている単語が失礼単語となる. システムのユーザは, この出力を確認して, 実際にメッセージとして投稿するか否かを判断することに役立てる.

表 4: 追加した失礼単語 (一部抜粋)

追加した失礼単語
ウザい キモい デブ
ケチ ショボい ダサい DQN
ニート 変態 ゆとり 無職

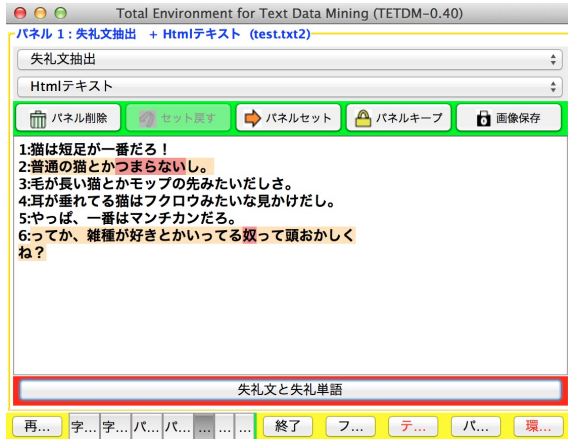


図 2: システムの出力例

## 4 システムの評価実験

本章では、システムがどの程度失礼文の抽出に効果があるかを検証した評価実験について述べる。

### 4.1 実験内容

提案するシステムによって、失礼文をどの程度抽出できるのかを明らかにするために、実験を行った。被験者は、情報科学を専攻する大学生・大学院生の計 20 名とした。表 5 に示す電子掲示板「2ちゃんねる<sup>2</sup>」のスレッドと、表 6 に示すニュース記事に対して、投稿された Twitter のツイートからそれぞれ 400 文、計 800 文を用いた。

手順 1 800 文全てを読んでもらう。この時、自分、誰か、もしくは何かに対し、失礼となる文を選択。(想定所要時間:40 分)

手順 2 システムが提示する失礼文抽出結果を参考に、手順 1 での回答を修正。(想定所要時間:20 分)

手順 2 では「手順 1 では選択したが、手順 2 では色付けされていない文」と「手順 1 では選択してなかったが、手順 2 では色付けされている文」の 2 パターンを修正の対象とした。また、被験者 20 名を 10 名ずつ、提案グループと比較グループの 2 グループに

<sup>2</sup><http://www.2ch.net>

分けた。提案グループには、失礼単語辞書(以下、提案辞書と記述する)を用いた失礼文抽出の結果を提示した。比較グループには、単語感情極性対応表の感情極性値が-0.48 以下となる 14035 単語を用いて作成した辞書(以下、比較辞書と記述する)を用いた失礼文抽出の結果を提示した。

表 5: 実験に使用した文を収集した電子掲示板のスレッド一覧

★鉱物・石ヲタスレ★ Vol.35
ネコさんと一緒に布団で添い寝したい PART11
これだけは理解できないってある? 2
100 円ショップの園芸モノってどうよ? その 23
東京の食いもんはマズい! マズすぎる! 166
【アイス総合】アイス大好きっ子スレ★その 97
チェリオのブルースコーヒーが不味すぎて困る
調理師として生きるということ part26

表 6: 実験に使用したツイートが投稿されているニュース記事

世界でもっとも任天堂グッズを集めた男がギネスに認定
【女性編】こんな人とカラオケに行きたくない! ランキング
イチロー選手のバッド窃盗容疑 神戸の公園、19 歳逮捕
当てはまったら要注意! 「パソコン依存症」な人の特徴 5 つ
任天堂「wii U」が救世主となれない 3 つの理由
「民間だったら当たり前」は「民間のブラック企業なら当たり前」
かつて『ハリー・ポッター』作者ももらっていた「生活保護」
ビンタ擁護論
「これで一斉に廃止したらどうなっちゃうのか」

### 4.2 実験結果と考察

表 7 に、提案辞書と比較辞書をそれぞれ用いた時のシステムが抽出した失礼文数と、各辞書の適合率、再現率を示す。正解文は、手順 2 の時点で、被験者 20 名中 10 名以上に失礼であると回答された 186 文とした。

表 7 より、提案辞書、比較辞書それぞれを用いたシステムの再現率の間に差はないが、適合率では大きな差がみられた。理由として、提案辞書を用いたシステ

表 7: 各システムの適合率と再現率

	適合率	再現率
提案システム	0.51(156/306)	0.84(156/186)
比較システム	0.25(170/677)	0.91(170/186)

表 8: 両グループ被験者の適合率, 再現率

	適合率	再現率
提案:手順 1	0.75(101/135)	0.54(101/186)
手順 2 後	0.74(139/188)	0.75(139/186)
手順 2 のみ	0.72(39/54)	—
比較:手順 1	0.69(105/153)	0.56(105/186)
手順 2 後	0.61(142/234)	0.76(142/186)
手順 2 のみ	0.46(37/80)	—

ムの総抽出文数が比較辞書を用いたシステムの出力文数より少なかったことと、提案辞書が少ない出力文数に対し、失礼文を多く抽出できていたことの2つが挙げられる。辞書の単語数によって、提案辞書の出力文数が、比較辞書の出力文数よりも少なくなったと思われる。これに加え、出力中の正解失礼文の数を見ると、提案辞書と比較辞書との間に大きな差がみられない。

以上より、提案辞書は比較辞書に比べ、単語数は少ないが、出力文数を抑えつつ、総正解文数 186 文中から 156 文、約 8 割の失礼文を抽出できたといえる。従って、提案する失礼単語辞書は、失礼文の抽出に有効な単語を収集されており、提案する失礼単語辞書による失礼文の抽出が、一定の精度で行えているといえる。

表 8 に、提案グループと比較グループの被験者の適合率、再現率の平均を、表 9 に提案グループと比較グループの修正時に被験者が追加した回答数とその中で正解だった回答数、また削除した回答数とその中で不正解だった回答数の平均を示す。表 8 より、提案、比較ともに、手順 2 後の再現率はほぼ同じ値となっているが、システムの出力を参考にした手順 2 における適合率には差があり、比較グループの被験者は非常に多くの失礼ではないと考えられる文も、システムの影響を受けて抽出していたことがわかる。また、表 9 より、提案は比較と比べて、修正時に適切な回答の追加と回答の削除を行えていることがわかる。このことから、提案辞書を用いたシステムにより、効率的に失礼文を抽出できることがわかった。

## 5 結論

電子掲示板や Twitter 等、多人数に向けメッセージから、失礼表現が含まれる可能性が高い文を抽出する

表 9: 両グループ被験者の回答追加, 回答削除数

	追加数	正解数	削除数	不正解数
提案	60	40	6	6
比較	84	38	4	3

システムを提案した。また、失礼表現が含まれる可能性が高い文を抽出する方法として、文に含まれる失礼表現となる可能性の高い単語に着目し、その単語を収集した失礼単語辞書を作成した。作成した失礼単語辞書による、失礼表現が含まれた文を抽出が、有効なことを明らかにした。

## 参考文献

- [1] 大坊郁夫: ネットワーク・コミュニケーションにおける対人関係の特徴, 対人社会心理学研究, No.2, pp. 1-14 (2002)
- [2] 濱岡秀平, 砂山渡: 単語特性辞書を用いた電子掲示板の雰囲気の同定, 日本知能情報ファジィ学会誌, Vol.24, No.3, pp. 707-716 (2012)
- [3] 一藤裕, 今野将, 曾根秀昭: 電子掲示板の雰囲気を考慮する発言分類, 電子情報通信学会技術研究報告, Vol.109, No.438, pp. 125-128 (2010)
- [4] 松田有史, 今井倫太, 大澤博隆: Twitter タイムライン解析による存在感の抽出, 全国大会講演論文集 2011, No.1, pp. 169-171 (2011)
- [5] 奥川巧, 大石哲也, 長谷川隆三, 藤田博, 越村三幸, 倉門浩二: Twitter のリスト機能を用いたユーザーの特徴抽出, 全国大会講演論文集 2011, No.1, pp. 687-689 (2011)
- [6] 古瀬蔵, 廣島伸章, 山田節夫, 片岡良治: ブログ記事からの意見文検索, 情報処理学会研究報告, 自然言語処理研究会報告 2006, No.124, pp. 121-128 (2006)
- [7] 池田和史, 柳原正, 松本一則, 滝嶋康弘: 係り受け関係に基づく違法・有害情報の高精度検出方式の提案, 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM2010), C9-5 (2010)
- [8] 石坂達也, 山本和英: Web 上の誹謗中傷を表す文の自動検出, 言語処理学会第 17 回年次大会 発表論文集, pp. 131-134 (2011)
- [9] 坂井俊之, 藤村孝: ブログに記述された不満表現からの潜在ニーズの発見, 情報処理学会論文誌, Vol.52, No.12, pp. 3806-3816 (2011)

- [10] 松葉達明, 里見尚宏, 榊井文人, 井須尚紀: 学校非公式サイトにおける有害情報検出, 電子情報通信学会技術研究報告, *NLC*, 言語理解とコミュニケーション, Vol.109, No.142, pp. 93-98 (2009)
- [11] 高村大也, 乾孝司, 奥村学: スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, Vol.47, No.2, pp. 627-637 (2006)
- [12] 松本裕治, 山下達雄, 平野義隆, 松田寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶筌』, Ver.2.4.0, 使用説明書 (2007)