Vol.2018 No.9,B1-2
2019/2/21

ブートストラップ手法に基づく Twitter 上のいじめ文の収集といじめ単語の抽出

橋向 慎[†1] 張 建偉[†1]

ネット上のいじめは重大な社会問題となっており、その自動検出には機械学習が効果的である。しかし、機械学習に必要ないじめ文といったものは現状公開されているものはない。本研究では、Twitter を対象として、いじめテキストを含むツイート文の収集と、いじめ文からいじめ単語の抽出を目的とする。収集手法としてブートストラップ手法を用い、基本単語を用いていじめ文の収集といじめ単語の抽出を行う。評価実験の結果、少数の単語と文章から大量のいじめ単語といじめ文を取得できる可能性を示した。

1. はじめに

現代社会は高度情報通信社会と呼ばれ、パソコンやスマートフォンから気軽にインターネットに接続できる。その一方で、ネット上でのいじめが年々増え続けている。平成 29 年の文部科学省の調査では、小・中・高等学校及び特別支援学校でのいじめの認知件数は 414,378 件となっている。そのうちネット上でのいじめは 12,632 件と全体の 3%に及ぶ。ネット上でのいじめは現実で行われるいじめとは違い、外部から見えにくく匿名性が高いため、発見されにくい性質があり、発見されていないいじめが存在する。日本の研究では、掲示板でのいじめの検出についての研究は行われてきている[1)2)3)]。

しかし、近年のネットいじめは掲示板だけでなく Twitter といった SNS 上でも発見されている[4)5)]。SNS 上のネットいじめを発見には、機械学習が有効的であり、学習に必要ないじめ文の収集は重要である。しかし、機械学習に使用できるような日本語のいじめ文は、現状公開されていない。Twitter 上のいじめの検出のために、いじめ文に用いられるキーワードを分析する必要性がある。

本研究では、Twitter を対象として、いじめテキストを含むツイート文の収集と分析を目的とする。Twitter 上ではデータが膨大なため、いじめテキストの割合が低い。そのため、いじめ文を効率よく収集することを目指す。収集手法として、ブートストラップ手法を用いる。まず始めに、Twitter 上から検索キーワードによるツイート文の収集を行う。収集した文がいじめかどうかは、クラウドワーカーに依頼することにより判断する。依頼にかかるコストを軽減するためにスコアを付与し、いじめの可能性が高いツイート文をランキングする。その後、いじめ文からいじめに使われる可能性の高い単語を抽出する。また、単語を抽出する手法に関しては、複数手法を検討する。

†1 岩手大学
Iwate University

2. 関連研究

海外では、Twitter 上での攻撃的なデータ収集は行われている。Waseem ら[4)] は、人種差別的な発言や性的暴行や差別的な発言を取得している。彼らは、ブートストラップ手法を用いての収集を試みている。宗教的や性的、少数民族に関して使われる単語を用いてツイート文と、これらの単語とともに出てくるハッシュタグ、攻撃的な発言の多いユーザの情報を取得する。これら 3 つから得られた情報からツイートを検索するという手順を繰り返し行うことにより、人種差別的な発言や性的暴行差別的な発言を取得している。Ross ら[5)] は、ヨーロッパ難民に対する攻撃的な発言の収集をしている。侮辱的または攻撃的な方法で使用されるハッシュタグを利用してツイート文を収集し、ツイート間の類似度に注目して攻撃的な文を収集している。

ネット上の悪口に関する研究例はこれまでにも複数存在する。松葉ら[6)] は、学校非公式サイトでの有害書き込みの検出をしている。彼らは、まず曖昧である有害情報の定義を Cohen の Kappa 値を用いて明らかにし、SVM によって有害情報の分類実験を行っている。石坂ら[1)] は、巨大電子掲示板「 2 ちゃんねる」を対象とし、誹謗中傷文を収集している。彼らは、単語悪口度の算出と悪口文/非悪口文の文分類の 2 つにより収集を試みている。事前に 2 つの基本単語を用意し, 対象の単語がその 2 つのどちらと文書内共起しやすいかを測定し、単語悪口度を算出している。その後、SVM を用いて悪口文と非悪口文に分類している。新田ら[2)] は、学校非公式掲示板を対象とし、有害書き込みの検出をしている。彼らは、有害書き込みのフレーズ抽出、有害語検出とカテゴリ化、関連度最大化による有害極性判定という三つの処理から書き込みを検出している。畠山ら[3)] も同様に、学校非公式掲示板を対象とし有害単語の抽出を検討している。彼らは、石坂ら[1)] の手法と新田ら[2)] の手法を組み合わせての抽出を行い、検討を行っている。李ら[7)] は、Youtube におけるコメント間のいじめ

検出をしている。彼らは、コメントの対象とコメント間の親子関係に着目し、辞書ベースでは検知できないいじめコメントの抽出に成功している。

いじめに限らず、ネット上のデータ収集手法の改善に関する研究も行われている。片岡ら[8] は、WEB 検索の際にユーザーが欲しい情報が得られるようなシステムの開発をしている。彼らは、検索結果のテキスト情報を要約し、検索キーワードと適合する WEB サイトのみを提示するシステムを開発している。田中ら[9] は、Twitter のツイートを基に、有用なユーザの推定している。彼らはユーザーのいいねやお気に入りの情報を用いて、ネットワーク解析による実験考察を行っている。松村ら[10] は、Twitter を対象に街に着目したツイート文の自動収集と分析システムの構築に挑戦している。彼らは、対象キーワードの一定時間の出現回数に注目し、急激に増えた場合その付近のツイート文を提示するシステムを構築している。また、水口ら[11] は、ブートストラップによる Web ページ上での辞書増語増殖に挑戦している。Web ページを効率よく収集することにより、少量のページ数でも辞書を構築でき、パターンを文字列で表現し単純化したことで容易に多言語の辞書を構築を可能にしている。

これまでにもネット上のいじめに関して研究は行われてきたが、日本語の Twitter 上のいじめに対して検出についての研究は行われていない。また、それらに対する分析や文章の収集も行われていない。著者が知る限りでは、日本語の Twitter 上のいじめに対しての研究は、本研究が初の試みである。

3. 研究手法

本研究では、Twitter 上からいじめ文の収集といじめ単語の抽出を行う。Twitter とは SNS の一種である。利用者数は平成 30 年 1 月現在、国内では約 4,500 万人、世界全体では 3 億人を超えており、利用者の多い SNS である。主な機能として、1 回につき 250 文字までのテキスト（ツイート）を投稿できる。研究手法として、ブートストラップ手法に基づき以下の順に行うことを提案する。ブートストラップ手法とは、少数の単語リストと文書からパターンを自動作成し、このパターンを使って文書から単語を抽出し、抽出した単語を使ってさらにパターンを自動作成する手法である。このような処理によって、少数の入力の単語リストを雪だるま式に大量に増やすことができる。また、いじめ文の収集やいじめ単語の中秋を効率良く行えることが予測される。以下に提案手法の手順を示す。

Step 1:Twitter API を用いてツイート文の検索を行う。Twitter API とは、Twitter 社が

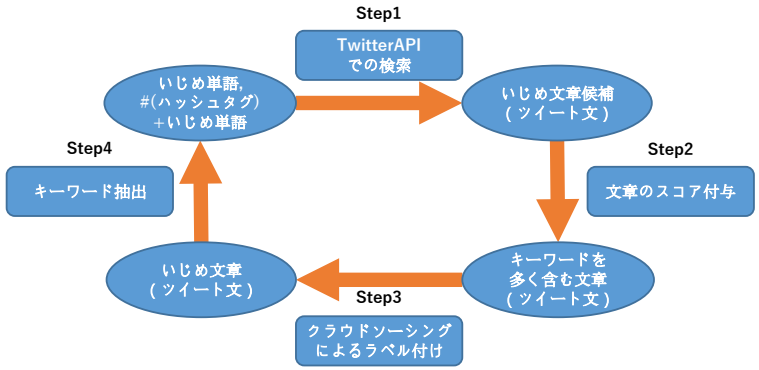

図 1 提案手法の流れ

公開している、Twitter データにプログラムレベルでアクセス可能なサービスであり、ツイートやユーザーの情報といった、様々な情報が取得可能である。いじめテキストが含まれているツイートを収集するために、基本いじめ単語として石坂ら[1] と新田ら[2] と畠山ら[3] の研究で用いられていたいじめ単語、合計 36 単語を使用する。しかし、Twitter では単語だけでなく、#（ハッシュタグ）+単語を使い文章を投稿する場合がある。ハッシュタグとは文章の検索を容易にするために投稿者が使用するタグ付け機能である。そのため検索キーワードとして 36 単語だけでなく、#いじめ基本単語も使用する。よって、検索キーワードは 36+36 = 72 単語を用いる。

Step 2:いじめ文章候補に対し、スコアを付与しランキングを行う。次の手順で行うクラウドワーカーによる判断件数を減らすために、基本いじめ単語が含まれている種類を基にし、ランキングを行う。

Step 3:クラウドソーシングによるラベル付けを行う。ランキング上位のツイート文に対しクラウドソーシングを利用し、いじめかいじめでないかのラベル付けを行う。

Step 4:クラウドワーカーによりいじめ文と判断されたツイート文から、新いじめ単語を抽出する。前処理として MeCab を用いて形態素解析を行う。形態素解析の結果に対し出現頻度や tf-idf、非いじめ文との比較を行い、新いじめ単語を抽出する。

これらの手順をブートストラップ手法として複数回行うことにより、Twitter 上からいじめ文を効率よく収集することができると考える。提案手法の流れを図 1 に示す。

3.1 ツイート文の収集

ツイート文を収集するために Twitter API を用いる。さきほども述べたように、検索キーワードとしていじめ基本単語 36 単語と#いじめ基本単語 36 単語、合計 72 単語を用いる。データとして、ツイート文だけでなく、投稿日時、ツイート ID、いいね数、リツイート数、ハッシュタグ数を取得する。ツイート ID とは、ツイートごとに振り分けられる 18 桁の数字で、数字が大きいほど最新のツイートとなる。いいねとは、投稿した内容に対し気軽に共感を伝えることができる機能である。しかし、気軽に使える機能のため、共感度は人によって異なる場合があったり、共感できない場合にも使われたりする。リツイートとは、最初に発信したツイートがほかのユーザーへ次々に拡散されていく機能である。この機能があるため、Twitter は Facebook や Instagram といった他の SNS と比べるとより拡散性が高い。

また、ツイート文にはデータの収集の障害となる文字が存在する。そのため、絵文字に関しては消去、ツイート文末の改行文字は消去、文末以外の改行文字やタブに関しては半角スペースに置換した。なお、これらの中には重複するツイート文があるため、ツイート ID を活用し重複しているツイート文を削除した。

3.2 ツイート文のスコア付与

この時点で収集できているツイート文は膨大である。そのため、クラウドワーカーによる判断件数を減らすためにツイート文にスコアを付与しランキングを行う。スコアは、文中に検索キーワードが何種類含まれているかで付与を行う。本研究では、いじめ基本単語と#いじめ基本単語を用いてのキーワード検索を行ったため、いじめ基本単語と#いじめ基本単語それぞれに重み付けをし、スコア付与を行う。文中に含まれるいじめ基本単語数を X、#いじめ基本単語数を Y とすると、ツイート文のスコア Z は、

$$Z = a * X + bY (\text{ただし、a+b=1})$$

となる。今回は重みを 0.1 刻みで変化させたため、11 種類の重み付けとなった。それぞれの重み付けに対し、スコア上位のツイート文数が 500 件以上になるようにツイート文を分類した。ただし、スコアの順位が 500 位のツイート文と同スコアのツイート文が 500 位以降もある場合は、同スコアのものも上位文とした。しかし、重み付けが適切かどうか判断できないため、一つの重みにつき、100 ツイート文ごとにランダムに 10 ツイート文選び、著者がいじめかどうか判断した。その割合が一定以上の重み付けの上位文を、Yahoo!クラウドソーシングに提出した。

3.3 Yahoo!クラウドソーシングによるラベル付け

スコアを付与した結果、上位のツイート文をいじめキーワードを多く含む文章とした。これらツイート文を Yahoo!クラウドソーシングを利用して、クラウドワーカーにいじめ・いじめでないのラベル付けをしてもらう。Yahoo!クラウドソーシングとは、Yahoo 社が提供するサービスの一種である。サービス内容として、企業や個人の大量のタスクを多くのサービス利用者に依頼できる場を提供している。他のクラウドソーシングプラットフォームとの違いとして、依頼内容に関して Yahoo!と相談できる点、実施したタスクの結果を tsv ファイルとして受け取ることができる点がある。人間が文をいじめかどうか判断する基準は様々であるため、大まかな基準ではあるが、著者側で分類基準を定義した。分類基準として、人に対して発言をしている前提で、攻撃的な文章である、差別的な文章である、個人情報を晒しているの 3 つを定義した。ただし、これら以外にも基準となる可能性があるため、その他のいじめ基準として判別した場合はチェック欄として「その他」を選んでもらい、自由記述欄に判断基準を書いてもらった。よって、いじめと判断した場合は分類基準を 3 つか「その他」をチェックしてもらい、いじめでないと判断した場合はいじめでないにチェックしてもらう。Yahoo!クラウドソーシングでの依頼画面と回答画面は図 2 のようになった。

図 2 依頼画面と回答画面

3.4 新いじめ単語の抽出

ラベル付けにより、いじめ文であるとしたツイート文に対し、MeCab を用いて形態素解析を行い、形態素ごとにする。ただし、不必要な単語を除外するため形態素として出力する品詞は、名詞、動詞、形容詞のみとした。この 3 つの品詞を基に、出現頻度や tf-idf を用いて新いじめ単語を抽出する。MeCab とは、京都大学情報学研究科と日本電信電話株式会社コミュニケーション科学基礎研究所の共同研究ユニットプロジェクトを通じて開発されたオープンソース形態素解析エンジンである。機能として、形態素解析、分かち書き、未知語の品詞推定がある。今回は 4 つの手法でいじめ単語の抽出を試みる。

3.4.1 出現頻度に基づく抽出

形態素解析の結果、残った単語に対し、一つの単語につき何回出現しているか数える。出現頻度が多い上位のものはいじめ単語やいじめに関連する単語である可能性がある。

3.4.2 tf-idf に基づく抽出

tf-idf とは、文書中に含まれる単語の重要度を評価する手法の 1 つであり、主に情報検索やトピック分析などの分野で用いられている。tf-idf は、単語の出現頻度：tf（Term Frequency）と逆文書頻度：idf（Inverse Document Frequency）の二つの指標に基づいて計算される。

$$tf\text{-}idf_{ij} = tf_{ij} * idf_i$$

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}$$

$$idf_i = \log \frac{|D|}{|d : d \ni t_i| + 1} + 1$$

n_{ij} は文書 d_j における単語 t_i の出現回数、$\sum_k n_{kj}$ は文書 d_j におけるすべての単語の出現回数の和、$|D|$ は総文章数、$|d : d \ni t_i|$ は単語 t_i を含む文章数である。そのため、idf は一種の一般語フィルタとして働き、多くの文書に出現する語は重要度が下がり、特定の文書にしか出現しない単語の重要度を上げる役割を果たす。

形態素解析の結果、残った単語に対し,tf-idf を使用し、ツイート文ごとの単語の tf-idf 値を求める。次に、単語ごとに tf-idf 値の合計を求める。tf-idf 値の合計値が大きいほど、いじめ単語やいじめに関連する単語である可能性がある。

3.4.3 出現頻度に基づくいじめ文と非いじめ文の比較による抽出

いじめ文の中には、日常的に使われる単語が含まれているため、ランキングした際にいじめ単語の割合が低くなる可能性がある。よって、ここまでに抽出された単語に対し、ランダムに収集した非いじめツイート文との比較を行う。これにより、日常的に使われる単語を削除でき、より精度良くいじめ単語を抽出することができる。ランダムに収集された非いじめ文に対し、いじめ文と同様に形態素解析を行い、出現回数を数える。次に、いじめ文に対する出現頻度に基づく抽出の結果から、ランダムに抽出した非いじめ文に対する出現頻度に基づく抽出の結果を削除する。

3.4.4 tf-idf に基づくいじめ文と非いじめ文との比較による抽出

ランダムに抽出した非いじめ文に対し tf-idf に基づく抽出を行う。いじめ文に対する tf-idf に基づく抽出の結果から、ランダムに抽出した非いじめ文に対する tf-idf に基づく抽出の結果を削除する。

4. 評 価 実 験

評価実験として、提案手法を用いて繰り返し処理の手順を 1 回行った。その結果と考察を以下に示す。

4.1 結　　果

今回の実験では、基本単語での検索では 2,282,501 ツイート、ハッシュタグ+基本単語での検索では 66,551 ツイート、合計 2,349,052 ツイートが収集できた。重複しているツイート文を削除した結果、1,733,779 ツイートがいじめテキストを含むツイート文候補として収集された。使用した基本いじめ単語と、単語それぞれの基本いじめ単語と#基本いじめ単語の検索数を表 1 に示す。

いじめツイート文候補にスコアを付与した結果、重み a が 1.0〜0.5、重み b が 0.0〜0.5 の値を取る場合のツイート文を使用した。使用するツイート文の重複部分を除くと、3,450 ツイートがいじめテキストを含むツイート文となった。また、著者の 100 ツイート文の収集ごとにランダムに 10 件ツイートを選択しいじめかどうか判断した際の結果は表 2 のようになった。

表 1 いじめ基本単語と検索数

ID	検索キーワード	いじめ単語	#いじめ単語	合計
1	消えろ	16,959	8	16,967
2	死ねよ	28,616	2	28,618
3	無能	52,885	55	52,940
4	ブサイク	24,201	75	24,276
5	ブス	156,783	415	157,198
6	童貞	144,172	8,069	152,241
7	チョン	4,694	18	4,712
8	死ね	156,933	110	157,043
9	ウザイ	15,765	9	15,774
10	キモイ	29,721	34	29,755
11	カス	104,642	61	104,703
12	殺す	116,518	4	116,522
13	クズ	153,947	709	154,656
14	うざい	28,049	45	28,094
15	キモい	47,319	90	47,409
16	気違い	1,731	4	1,735
17	ウジ虫	1,988	0	1,988
18	嫌い	718,874	107	718,981
19	ックス	238,265	2,1016	259,281
20	ヤリマン	4,761	2,838	7,599
21	フェラ	60,916	27,345	88,261
22	殴る	41,373	2	41,375
23	ダセー	666	0	666
24	馬鹿だよ	36,574	0	36,574
25	糞尿	1,649	1	1,650
26	マジかも	17,595	0	17,595
27	イボオタ	0	0	0
28	クズマスゴミ	15	0	15
29	ゴキオタ	0	0	0
30	ビッチ	30,935	5,382	36,317
31	目くそ	1,113	16	1,129
32	脱糞	9,220	36	9,256
33	糞虫	12,076	2	12,078
34	マジキモ	571	1	572
35	愚民	10,899	38	10,937
36	寄生虫	12,076	59	12,135
		2282501	66,551	2,349,052

表 2 重み付けと判断結果

重み a/b	文数	スコア 最大/最小	著者が判断した文数	いじめと判別した文数	割合
1.0/0.0	**2,737**	**6.0/3.0**	**280**	**39**	**0.139**
0.9/0.1	**2,737**	**5.4/2.7**	**280**	**36**	**0.129**
0.8/0.2	**2,737**	**4.8/2.4**	**280**	**32**	**0.114**
0.7/0.3	**2,765**	**4.2/2.0**	**280**	**47**	**0.168**
0.6/0.4	**2,769**	**3.6/1.8**	**277**	**39**	**0.141**
0.5/0.5	**3,450**	**3.0/1.5**	**346**	**53**	**0.153**
0.4/0.6	922	2.6/1.4	93	6	0.065
0.3/0.7	4,580	2.7/1.4	460	0	0.000
0.2/0.8	4,551	2.8/1.6	460	0	0.000
0.1/0.9	4,551	2.9/1.8	460	1	0.002
0.0/1.0	4,551	3.0/2.0	460	2	0.004

表 3 アンケート結果の集計

いじめと判断した人数	文章数	攻撃的な発言をしている	差別的な発言をしている	個人情報を晒している	その他
3	1,395	3,221	1,712	499	409
2	1,013	1,415	869	294	326
1	760	408	351	189	263
0	282	31	23	16	52
合計	3,450	5,075	2,995	998	1,050

Yahoo!クラウドソーシングを利用した結果、1,395 ツイート文がクラウドワーカーが 3 人一致でいじめであると判断したツイート文、1,013 ツイート文が 2 人一致でいじめであると判断したツイート文、760 ツイート文がクラウドワーカーが 1 人だけいじめであると判断したツイート文、282 ツイート文が 3 人一致でいじめでないと判断したツイート文となった。一人に判断してもらうツイート文は 10 ツイート文とし、一つのツイート文に対し、3 人で評価してもらった。今回は一つのツイート文に対し、3 人一致でいじめであると判断したものをいじめ文であるとした。アンケートの回答結果の集計を表 3 に示す。

また、クラウドワーカーの回答のばらつきを判断するために、Fleiss の κ 係数を用いてクラウドワーカー間の評価の一致度の評価を行った。対象は第 3.3 節で行ったアンケートとした。Fleiss の κ 係数とは、評価者間の一致の信頼性を評価するための統計的尺度であり、2 人以上の評価者の間の一致を評価する場合に有効である。アンケート対象のツイート文数を N、一つのツイート文の回答者数を n、アンケートの回答の分類を k、それぞれのツイート文を i、それぞれのアンケートの回答の分類を j とすると、

係数: $\kappa = \dfrac{\overline{P} - \overline{P_e}}{1 - \overline{P_e}}$

観察した一致率: $\overline{P} = \dfrac{1}{Nn(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{k} (n_{ij}^2 - Nn)$

偶然の一致率: $\overline{P_e} = \sum_{j=1}^{k} (\dfrac{1}{Nn} \sum_{i=1}^{N} n_{ij})^2$

で表わされる。結果、κ=0.22 となった。κ 値が 0 以下の場合、一致しなかったということになるが、0.22 の場合、ある程度の一致となったことになる。κ 値の目安を表 4 に示す。

表 4 κ 値の目安

κ	解釈
0 未満	一致しない
0.0〜0.2	わずかに一致
0.21〜0.40	ある程度の一致
0.41〜0.60	中等度の一致
0.61〜0.80	かなりの一致
0.81〜1.0	ほぼ完全な一致

出現回数上位 100 単語に対し、著者がいじめに関する単語かどうか判断した結果を表 5 と図 3 に示す。上位 N 単語 (N=10,20, ⋯ ,100) について、表 5 はいじめ単語が何個含まれているか、図 3 はいじめ単語が含まれる割合を表している。

表 5 いじめに関する単語の個数

	出現頻度に基づく抽出	tfidf に基づく抽出	出現頻度に基づくいじめ文と非いじめ文の比較による抽出	tf-idf に基づくいじめ文と非いじめ文の比較による抽出
N=10	0	0	7	7
N=20	3	5	12	13
N=30	4	7	14	18
N=40	5	9	20	24
N=50	8	12	26	30
N=60	8	12	31	36
N=70	11	16	36	43
N=80	13	18	40	47
N=90	17	22	44	53
N=100	20	25	49	54

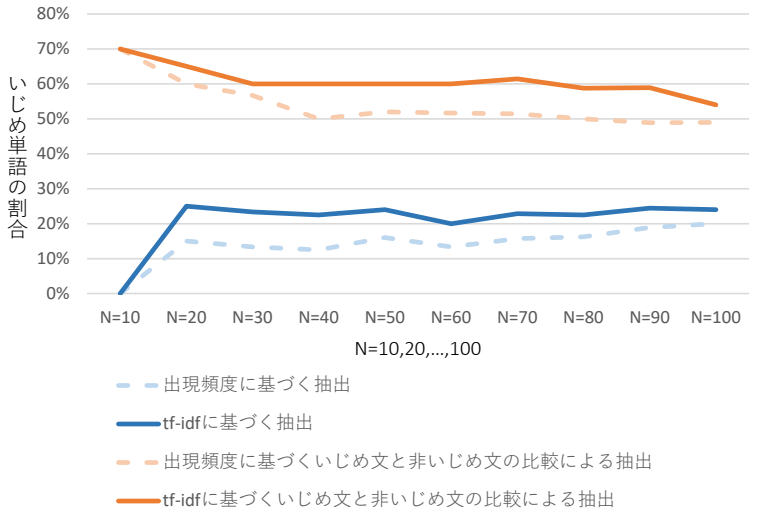

図 3 いじめに関する単語の割合のグラフ

表 6 抽出手法によって出現した単語

	出現頻度に基づく抽出	tfidf に基づく抽出	出現頻度に基づくいじめ文と非いじめ文の比較による抽出	tf-idf に基づくいじめ文と非いじめ文の比較による抽出
1 位	お前	お前	嫌いな	嫌いな
2 位	ー	ー	糞	糞
3 位	人	人	きもい	きもい
4 位	な	こと	こいつ	こいつ
5 位	こと	な	キャ	キャ
6 位	じ	女	殺意	殺意
7 位	女	じ	くそ	馬鹿
8 位	私	私	馬鹿	くそ
9 位	マジ	マジ	キモ	キモ
10 位	の	の	共	共

いじめ文中の出現頻度や tf-idf を用いた単語の、非いじめ文との比較の結果、新いじめ単語として「デブス」や「ゴミクズ」、「臭い」のような単語が抽出できた。それぞれの抽出手法によって出現した単語上位 10 単語を表 6 に示す。

5. 考　　察

実験結果に対し、今回収集できたいじめ文、及び抽出できた新いじめ単語について考察する。収集できたいじめ文を見てみると、大部分のものが誰がみてもいじめ文だと判断できるものが出現した。しかし、一部のものは受け取り手によって意見が多く変わるものもあった。改善策として、ツイート間の結びつきもクラウドワーカーへ提示する必要があると考える。

今回のアンケートではクラウドワーカーが真剣に回答してくれるかどうかは考慮していない。そのため、今後の研究ではアンケートの中に数問、誰が見てもいじめかどうか判断できる文を追加しその正解率を見ることでクラウドワーカーの信頼度を判断する必要がある。

また、いじめの分類に関しても詳細を考えていく必要がある。図 3 を見てみると、いじめと判断した場合の大部分が攻撃的な発言をしているに回答していた。その他のいじめ基準として、性的な発言であるや、人道的な発言でないといった意見があった。今後の研究の際にこのような分類も考慮していく必要がある。

Fleiss の κ 係数は 0.22 という結果になった。アンケート結果がある程度の一致となった理由として、アンケートの分類をいじめかいじめかでないかの 2 種類で判断したため、偶然の一致率が高くなってしまったことが考えられる。

6. まとめと今後の課題

本研究ではブートストラップ手法に基づく Twitter 上のいじめ文の収集といじめ単語の抽出を行った。実験により、ブートストラップ手法によるいじめツイート文の収集及び新いじめ単語の抽出の可能性を示した。今後は今回抽出した新いじめ単語を使用してのいじめツイート文の収集、クラウドソーシングでのアンケート方法や解析手法の改善を行っていきたい。

参　考　文　献

1) 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, 2011.
2) 新田大征, 桝井文人, PtaszynskiMichal, 木村泰知, RzepkaRafal, 荒木健治. カテゴリ別関連度最大化手法に基づく学校非公式サイトの有害書込み検出. 第 27 回人工知能学会全国大会, 2013.
3) 畠山鈴生, 桝井文人, プタシンスキ ミハウ, 山本和英. 有害表現抽出に対する種単語の影響に関する一考察. 第 30 回人工知能学会全国大会, 2016.
4) Zeerak Waseem, and Dirk Hovy. Hateful Symbols or Hateful People?Predictive Features for Hate Speech Detection on Twitter. Proceedings of the NAACL Student Research Workshop, 2013.
5) Bjrn Ross, Michael Rist, Guillermo Carbonell, Benjamin Cabrera, Nils Kurowsky, and Michael Wojatzki. Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis. NLP4CMC III: 3rd Workshop on Natural Language Processing for Computer-Mediated Communication, 2016.
6) 松葉達明, 桝井文人, 河合敦夫, 井須尚紀. 学校非公式サイトにおける有害情報検出. 言語処理学会 第 16 回年次大会 発表論文集, 2010.
7) 李子怡, 川本淳平, フォンヤオカイ, 櫻井幸一. コメントの親子関係を利用した動画共有サイトにおけるネットいじめコメントの検出. DEIM Forum 2016, 2016.
8) 片岡充照, 今中武, 水谷研治, 若見昇. テキスト情報を対象としたキーワード抽出と関連情報収集システム. 日本ファジイ学会誌, Vol.9, No.5 pages:710-717, 1997.
9) 田中淳史, 田島敬史. twitter のツイートに関する分類手法の提案. DEIM Forum 2010, 2010.
10) 松村飛志, 安村通晃, 街に着目した Twitter メッセージの自動収集と分析システムの提案と試作. 情報処理学会 インタラクション, 2010.
11) 水口弘紀, 河合英紀, 土田正明 久寿居大. Web 知識を利用したブートストラップによる辞書増殖手法. 電子情報通信学会 第 18 回データ工学ワークショップ, 2007.