

ID 交換掲示板における書き込み有害性評価に向けた 隠語概念化手法の提案

† 山田 大 †† 安彦 智史 † 長谷川 大 ‡ Michal Ptaszynski †† 中村 健二 † 佐久田 博司

† 青山学院大学 情報テクノロジー学科 †† 青山学院大学附置 情報メディアセンター

‡ 北見工業大学 情報システム工学科 †† 大阪経済大学 情報社会学科

a5812113@aoyama.jp, abiko@aim.aoyama.ac.jp, {hasegawa,sakuta}@it.aoyama.ac.jp,
ptaszynski@cs.kitami-it.ac.jp, k-nakamu@osaka-ue.ac.jp

1 はじめに

ネットワークサービス利用者の低年齢化に伴い、青少年の安全なインターネット利用は重要な課題となっている。違法・有害な情報を含むサイトは増加傾向にあり、2013年にはID交換掲示板と呼ばれるサイトに起因する犯罪被害数が一年間で約10倍に増加した¹。

これまでの違法・有害情報に関連する研究では、サイト内の単語分布により有害判定を行う研究 [1] や有害サイトに特徴的に見られるHTMLを基に有害判定を行う研究 [2] がある。また、文書の特徴を用いて共起関係や係り受け関係によって有害情報を検出する研究 [3][4] がある。しかし、ID交換掲示板の書き込みには多様な隠語表現が含まれており、形態素に分割できず正確な有害判定が困難となっている。周辺語から隠語の意味を決定する研究 [5] もあるが、対象のような短文や形態素解析が困難な文書に対しては有効ではない。

そこで、本研究ではID交換掲示板における有害情報を含む書き込みの検出精度向上を目的として、書き込みに含まれる隠語表現を分類し、それぞれの隠語表現に対する概念化手法の提案を行う。

2 隠語表現の分類

本研究では、隠語表現の概念化を行うために6つの掲示板から収集した書き込み600件を対象に隠語表現を詳細に分類した。分類した隠語の種類を表1、書き込みに含まれる隠語表現の割合を図1に示す。ID交換掲示板で利用される隠語では、短縮・意味や表層的言い換えに分類されるものが多く、次いでマスク、当て字となった。言外の情報については出現頻度が少ない

表 1: 隠語表現の種類

	説明	具体例
短縮・意味	名詞や動詞を省略や言い換えをして表現	ぶちなどOKなん で相談下さい。
表層的 言い換え マスク	単語の間に記号などの文字列を含めて表現 単語の一部を記号などに置き換えて表現	えっ、、、、、、 ち、、、い え。ち
当て字	単語の一部を類似する他の文字に置き換えて表現	女の子ばかりく わえててずるい!
言外の 情報	直接的な単語を用いずに特定の意味を表す表現	意味わかる人

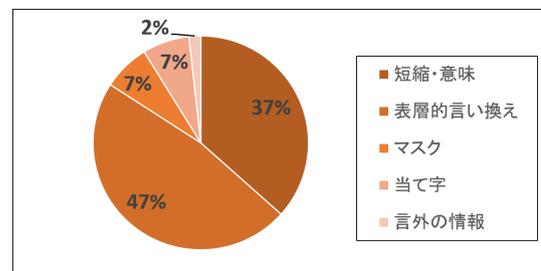


図 1: 隠語表現の割合

こと、既存研究でも対応可能なことを考慮し、短縮・意味、表層的言い換え、マスク、当て字を本研究の対象とした。

3 手法

本提案手法では、2章で行った隠語表現の分類に基づき、それぞれの隠語に対応した処理を施すことで、隠語を概念化したデータセットを作成する。処理の流れについては図2に示す。当て字変換処理部では、当て字辞書を用いて当て字として置き換えられた文字に対応する文字に変換する。次に、表層的言い換え変換

¹警視庁 <https://www.npa.go.jp/cyber/statics/h27/h27-1.pdf>

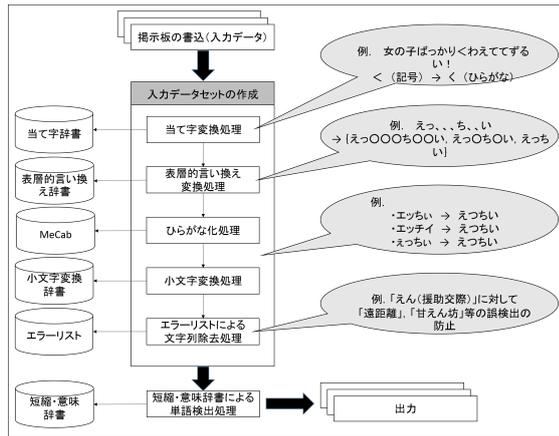


図 2: 処理の流れ

処理部では表層的言い換え辞書に対応する文字をマスク化し、原文・すべてをマスク化したもの・連続した記号をマスク化したもの・マスク化してないものなどでデータセットを作成する。ひらがな化処理部で MeCab² を用いて読みがなを取得し、小文字変換処理部で小文字を大文字へと変換する。エラーリストによる文字列除去処理部においては、有害語検出を行う際にひらがなで処理を行うことによる誤検出を防止するため、有害語と同一の文字列だが意味の異なったデータセット内の文字列を除去している。この順序で処理を行うことで表層的言い換え、マスク、当て字の隠語表現を含む書き込みに対して概念化したデータセットを作成し、最後に短縮・意味辞書を用いることで全ての隠語表現に対応する。各変換処理で用いた辞書は 3 つの ID 交換掲示板から収集した書き込み 300 件を学習データとして、隠語表現に用いられた単語や記号に対応する辞書へ登録した。

4 実験

4.1 実験方法

実証実験では、本研究における本提案システムの精度を評価するため、学習データとは異なる 3 つの ID 交換掲示板から収集した書き込み 300 件を対象に検証を行った。検証では、人手により判定した処理結果を正解データとし、本提案システムから出力した処理結果と比較することで、本提案システムの適合率と再現率、F 値を算出する。なお、書き込みにおける隠語表現の割合は短縮・意味 40%、表層的言い換え 42%、マスク 6%、当て字 12% であり、用いられた隠語表現の数は 132 件であった。

²MeCab <http://taku910.github.io/mecab/>

表 2: 評価実験における適合率、再現率、F 値

	短縮・意味	表層的 言い換え	マスク	当て字
適合率	0.627	1.000	1.000	1.000
再現率	0.698	0.945	0.875	0.875
F 値	0.661	0.972	0.933	0.933

4.2 実験結果と考察

実験結果より、各隠語表現の適合率、再現率、F 値を表 2 に示す。正解データに対して正しく変換された隠語表現は 110 件であり、これは隠語表現全体の 83% であった。表層的言い換え、マスク、当て字については、F 値 0.9 程度と高精度に処理を行えたが、短縮・意味については、0.6 程度となった。これは、ID 交換掲示板の書き込みに含まれている短縮・意味に該当する表現が多様であり、辞書による単語のみでは対応しきれなかったことが原因だと考えられる。表層的言い換え、マスク、当て字については、用いられる文字の種類が限定されていたため高い精度を示したと考えられる。

5 おわりに

本研究では、ID 交換掲示板における書き込みに含まれる多様な隠語表現を分類し、隠語表現の概念化を行う手法を提案した。

今後の展望として、短縮・意味に対する辞書データを充実させることで、より高い精度を目指す。また、提案手法を用いて作成したデータセットを機械学習させることで、既存手法では判定できなかった隠語表現を含んだ書き込みに対してより精度の高い有害性評価を行うことが出来るようになる。

参考文献

- [1] 井ノ上直己, 帆足啓一郎, 橋本和夫. 文書自動分類手法を用いた有害情報フィルタリングソフトの開発. 電子情報通信学会論文誌, pp. 1158-1166. 電子情報通信学会, 2001.
- [2] 池田和史, 柳原正, 松本一則, 滝嶋康弘. 格要素の抽象化に基づく違法・有害文書検出手法の提案と評価. 情報処理学会第 72 回全国大会, pp. 71-72. 情報処理学会, 2010.
- [3] 中村健二, 田中成典, 山本雄平, 安彦智史. 共起関係の抽出範囲を考慮した有害情報フィルタリング手法. 情報処理学会, pp. 571-584, 2013.
- [4] 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. In *DEIM Forum 2010*. 日本データベース学会, 2010.
- [5] 橋本広美, 木下嵩基, 原田実. フィルタリングのための隠語の有害語意検出機能の意味解析システム sage への組み込み. Technical report, 情報処理学会, 2010.