卒業論文要旨

人権ネットパトロールシステムの効率向上に関する研究

（情報環境学）　久保田　剛史

1. 緒言

近年、インターネットの発達に伴い、掲示板やSNSといった情報発信システムへの個人情報の掲載によるプライバシーの侵害、特定の個人を対象とした誹謗・中傷や差別的な表現の書き込みなどが大きな問題となっている。本研究室では、代表的なSNSである「Twitter」を対象に、こうした問題のある書き込み（ツイート）を検出するシステムとして、人権ネットパトロールシステムの研究開発を進めている（例えば、文献[1]）。本研究では、ツイートに含まれる人名と人権侵害語の形態素間距離、および文書分類[2]を用いた代表ツイートの抽出による人権ネットパトロールシステムの効率向上法を提案する。

2. 実験環境

OS：Windows7 Professiona

PC：DELL Optiplex 3020

使用プログラミング言語：Microsoft Visual C# 2010,Python 2.7.3

3. 既存システムの課題

既存システム[1]ではツイートごとに MeCab[3]による形態素解析を行い、人名と人権侵害語が検出されたツイート等を通報候補と判定し、目視チェックに供する。しかし、既存システムでは通報候補として、特定人物が人権侵害語により誹謗・中傷を受けているツイートの他に、検出された人名の人物が人権侵害の対象になっていないツイートや、人権侵害語や人名でない形態素を誤って検出しているツイート（以下、過検出）も含まれているため、目視チェックに過剰な時間を要する。

3.1 人権侵害語に関する過検出と対策

特定の単語が組み合わさった時に、以下のように、誤って人権侵害語として検出している。

@○○○そっかーならよかたわ((

この例では、「かたわ」が人権侵害語としてMeCabの辞書に登録されているため、枠線で囲んだ範囲を優先して検出し、過検出となった。そこで、対策として、過検出の原因となっている単語（上例の場合「よかたわ」）をMeCabの辞書に登録することにより、過検出の発生が減少した。

3.2 人名に関する過検出

ツイート中で人名ではない単語を人名としてし検出した場合やツイート中に検出した人名の人物が人権侵害の対象になっていない場合があげられる。

4. 通報候補の人名と人権侵害語との形態素間距離に関する考察

人名と人権侵害語が同時に検出される通報候補から通報対象ツイートを見つけ出す作業の効率を向上させるために以下の考察を行った。なお、考察の際に昨年度に本研究室で検出した通報対象ツイート68件を利用した。

4.1 形態素間距離

それぞれのツイートに対してMeCabを用いて形態素解析を行い、人権侵害語と判定された形態素と、人名と判定された形態素との間に存在する形態素の数をそれぞれ数え、最少の形態素の数を「形態素間距離」として取扱った。

4.2 通報対象ツイートの形態素間距離

前述した通報対象ツイート 68 件の形態素間距離を図 1 に示す。

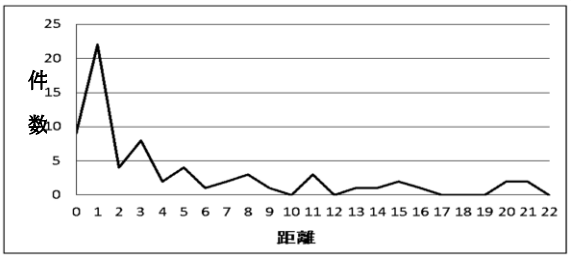

図 1. 通報対象の形態素間距離の度数分布

4.3 通報対象ツイートと通報不要ツイートの形態素間距離の比較

ツイートにおける人名と人権侵害語との形態素間距離を用いて通報候補から通報対象ツイートを効率良く見つけ出せるかどうかを検討するため、新たにツイートを収集した。Yahoo!リアルタイム検索を用いて、前述の通報対象の中で数が多かったキーワードである「在日」と「チョン」で検索を行い、既存システムで通報候補と判定されたツイートを 1194 件収集した。収集したツイートの内、実際に通報対象と判断したものは 8 件あった。これら 1194 件と通報対象ツイートの中に「在日」と「チョン」が含まれているツイート 38 件の形態素間距離を求めた。その結果、通報対象ツイートについては人名と人権侵害語の形態素間距離が 1 以下に 47.06%存在しているのに対し、通報不要ツイートについては形態素間距離が 1 以下に 26.21%存在していることが判明した。このことから、通報対象ツイートの存在割合が高い形態素間距離 1 以下のツイートだけを目視チェック対象とすることによって、人権ネットパトロールシステムの目視チェック時間を短縮できると考える。

5. 文書分類によるツイートのクラスタリングと代表ツイートの抽出

Twitter 上では話題となっているトピックスや、注目されているツイート等が多くの人に引用され、その結果、類似しているツイートが多数存在する。このため、これらのツイートに対して文書分類[2]を行うと、類似しているツイートは一つのグループにクラスタリングされ、クラスタの重心に最も近いツイートを代表ツイートとして抽出することができる。また、クラスタリングされたツイートの中で多くのツイートが含まれているクラスタを参照し、代表ツイートが人権を侵害する文章だった場合、そのクラスタ内のツイートに引用元となった発言、あるいは情報発信サイトの URL が発見できる可能性がある。発見できた情報発信元のブラックリスト化は人権ネットパトロールの効率向上につながると考えられる。

6. 結言

人名と人権侵害語の形態素間距離をもとに、一定値以下のツイートのみを目視チェックすることで作業の効率化を図れる可能性を見出した。しかし、形態素間距離の値が大きい通報候補について検出漏れが発生してしまうことが課題である。また、ツイートの文書分類に関しては、処理時間の短縮が今後の課題である。

[参考文献]

[1] 上田裕果,「人権ネットパトロールシステムに関する研究」, 平成 25 年度京都府立大学生命環境学部環境・情報科学科卒業論文.

[2] J.Kimura, Y.Yoshitomi, and M.Tabuse, “Classification of Japanese Documents and Ranking of Representative Documents Using Characteristic of Frequencies of Words”, Proc. of Int. Conf. on Artificial Life and Robotics, 2015, pp.306-309.

[3] 京都大学情報学研究科－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共同研究ユニットプロジェクト, http://mecab.sourceforge.net/