6X-03

児童被害を抑止するためのSNS上の不正コメント抽出方法

住田　淳[†]　亮　隆弘[†]　菱田　隆彰[†]
愛知工業大学[†]

1 はじめに

ソーシャルネットワーキングサービス（SNS）が広く普及した現在，SNSを利用した援助交際の呼びかけによる児童被害が急増している．

警察ではSNSの内容を元に補導などの被害防止活動を行なっているが，情報収集作業は他の職務の間の限られた時間に全て手作業で行われ，得られた情報の共有も困難であることが効率の悪化を招いている．

本研究では，その課題を解消するための手法として，機械学習を用いSNSで使用されるワードや隠語を元に危険と判断した投稿を抽出する手法を提案し，その情報を電子的に共有することで効率よく情報収集支援を可能とするシステムの試作を行う．

2 児童被害防止活動の課題

平成15年の出会い系サイト規制法施行や平成20年の出会い系サイト規制法改正に伴い出会い系サイトにおける被害児童数は年々減少し，平成28年では過去最少を記録している．他方，コミュニティサイトにおける被害児童数は増加傾向が継続し，平成28年では過去最多を記録している[1]．児童被害数が多いサイトとしてマイクロブログが挙げられ，全体の4割を占める．

警察ではマイクロブログで検索した内容を元に補導などの活動を行なっている．しかし，多くの投稿の中から援助交際に関する内容を見つけ，実際に補導のため児童に会う約束を取り付けるまでの作業を限られた少ない時間に手作業で行うため，全ての投稿を精査しきれないという課題がある．

マイクロブログの投稿内容の分析については様々な手法が用いられている．荒川ら[2]は，ランダムフォレスト機械学習法を用いて投稿の分類を行い，リツイート数と紐付けることで拡散されやすい投稿の特徴を検討している．川本ら[3]は，リツイート数により影響力を測定し，RBFカーネルのサポートベクタマシンを用いて投稿の分類をしている．

本研究では，機械学習を用いて投稿から補導の対象となる不正なコメントを抽出し，危険度別に分類を行うシステムを試作する．本システムを使用することで多くの投稿の中から不正コメントを検出する時間が短縮され，児童被害防止活動を効率化し，児童がより安全に過ごせる環境を実現できる．

3 マイクロブログ上の投稿の分類

3.1 不正コメント抽出方法

本研究では，ランダムフォレスト機械学習法を用いて分類を行う．不正コメントを抽出する方法として，文章を特徴語の出現頻度によってベクトル化し，その特徴ベクトルを分類器にかけて文章のカテゴライズを行う．モデルの学習を行うため，教師データを用意する．教師データを学習させることで収集した文章から不正コメントの抽出，危険度の分類を行う．分類に使用する特徴ベクトルの生成は，始めに文章を形態素解析し，単語リストに変換する．次に単語リストから特徴語の辞書を作成する．実際に作成した特徴語の辞書は22855個の成分を持つ．最後にBag-of-Wordsの要領で各文章に特徴語が何個含まれるのかを数え特徴ベクトルとする．

3.2 教師データの作成

教師データの作成には愛知県警の警察官に協力していただいた．表1に示すように，一ヶ月間取得した投稿は目視で確認し0から5の6段階評価のラベル付けがされている．今回取得した投稿のうち1割が不正コメントで

表 1:1ヶ月間に取得した投稿の内訳

種別		投稿数
全投稿		17834
不正コメントでない		16153
不正コメントである	カテゴリ0	1599
	カテゴリ1	2
	カテゴリ2	31
	カテゴリ3	22
	カテゴリ4	9
	カテゴリ5	18

A method of illegal comments extraction on social network sites to protect children from harm
Jun Sumita[1] ,Akira Takahiro[1] ,Takaaki Hishida[1]
[1]Aichi Institute of Technology

あった．評価ラベルに関しては，数値が上がるにつれ児童が被害を受ける危険度が増していく．カテゴリ 0 の内容は bot からの投稿など直接的な内容ではなく補導が難しいものとなる．カテゴリ 1 以降は協力していただいた警察官の主観に基づきラベリングがされており，投稿者の年齢や募集している時間，場所など投稿内容の具体性が上がるにつれて危険度が上がっている．また，カテゴリ 5 の内容は，年齢や募集場所の他に詳細な値段設定が書かれており，「今日の夕方」など比較的早く補導がかけられるような内容や，すぐにメッセージに反応してくれそうだと感じる内容で，緊急性の高いものである．

4 不正コメント抽出システム

4.1 システムの概要

前述の手法に従い不正コメント抽出システムの構築を行う．本システムはマイクロブログ上から過去に不正コメントで使用されてきたワードや隠語を用いて投稿の収集を行い，その内容に含まれる特徴語を分類し危険度別にカテゴライズする．投稿内容の分類した結果を警察と共有するために Google Spread Sheet を使用する．システムの動作の流れは，大きく分けて以下の３つの処理を行う．

1) 収集処理:マイクロブログ上からキーワード検索された投稿を REST API により取得する
2) 解析処理:不正コメントを抽出し危険度別にカテゴライズした結果を Sheet に書き込む
3) 共有処理：Sheet の内容を警察の用意したアカウントと共有し閲覧可能にする

これらの処理を毎日一回実施することで，担当の警察官は常に最新の情報を用いて活動できる．

4.2 投稿の収集処理

今回我々は, 情報収集の対象として Twitter を選んだ．Twitter は日本国内における月間利用者数 4000 万人と非常に多く, 被害児童数が全体の４割を占めることが選定の理由である．

本システムは, Twitter が用意している API を使用して投稿の検索，取得を行う．分類対象となる文章の取得は，不正コメントに使用されるワードや隠語を用いて行う．取得した投稿から補導するために必要な「ユーザーネーム」,「スクリーンネーム」,「プロフィール」,「投稿内容」,「投稿 URL」を抽出しシステム内のデータベースに保存する．

4.3 投稿の解析処理

先の収集処理によって得られた投稿内容を対象に分類を行う．今回使用した分類器は二つある．まず表 1 に示す教師データの不正コメントでない投稿 16153 件と不正コメントである投稿 1681 件を学習させた分類器を使用し，不正コメントであるかの分類を行う．次に不正コメントであるデータの各カテゴリを学習させた分類器を使用し，危険度別にカテゴライズを行う．各カテゴリにラベリングされた投稿を降順でソートし Sheet に書き込む．

4.4 投稿の共有処理

Sheet はクラウド上に保存されており，複数人のユーザーが同時にアクセスし，リアルタイムに書き込みを閲覧できる．投稿は毎日取得し解析するため，曜日毎の Sheet を作成する．書き込まれた Sheet は警察で用意したアカウントと共有を行う．Sheet を共有している警察官は上から順に補導をするためのアクションをかけることで効率的に児童被害防止活動を実施することができる．

5 まとめ

本研究では，Twitter を対象に児童被害の要因となる投稿を抽出し，危険度の高い投稿に関する情報を速やかに担当の警察官と情報共有するための手法の提案とそのシステムの試作を行なった．

教師データの 6 割を学習させ，4 割をテストデータとして分類した結果，不正コメントの抽出を行う分類器の正答率は 99%，危険度別にカテゴライズを行う分類器の正答率は 97%となった．本システムを利用することで，多大な時間や労力をかけることなく児童被害防止活動を行うことが可能となった．

今後の課題として，現在本システムの試験運用を行なっており，新しく取得し分類したデータの評価を行いたい．また，投稿者の性別も危険度に大きく影響を与えるため，文章から投稿者の性別を分類する機能の追加を検討したい．

謝辞

本研究は平成 29 年度愛知工業大学教育・研究特別助成の助成を受けて行なっている．

参考文献

[1] 平成 29 年上半期におけるコミュニティサイト等に起因する事犯の現状と対策について (http://www.npa.go.jp/cyber/statics/h29/H29_siryou.pdf) , 警察庁， アクセス日：2017/12/20

[2] 荒川唯，亀田尭宙，相沢彰子，鈴木崇史:Retweet に着目した広がりやすい Tweet の特徴分析, 情報処理学会第 74 回全国大会講演論文集，2012-1,pp.617-618,2012

[3] 川本貴史，豊田正史，吉永直樹:マイクロブログからの社会的影響力を持つ情報カスケードの検知手法，情報処理学会論文誌データベース，9-2,pp.23-33,2016