

Instagram を対象としたサイバーパトロールの効率化手法の検討

横山 未有[†] 角田 裕[†][†]東北工業大学 工学部情報通信工学科

1. はじめに

スマートフォン等の普及に伴い、Social Networking Service (以下、SNS) に起因する性被害を受ける児童が年々増加している [1]。対策として、SNS 上の有害な投稿を事業者や警察に通報するサイバーパトロールが行われているが、主な活動対象は Twitter であり、Instagram での取り組みは行われていない。また、サイバーパトロールでは限られた人数で多くの有害投稿を扱うこととなるので、効率化の取り組みが必要とされる。サイバーパトロールを効率化するための研究についても、Twitter を対象とした研究 [2]-[5] は多数あるが、Instagram を対象とした研究は、筆者の知る限り行われていない。

本研究では、今後も児童性被害の原因になりうると予想される Instagram において、有害投稿がどの程度存在するのかを確認し、サイバーパトロールの必要性を示す。また、Twitter に関する先行研究を参考に、Instagram におけるサイバーパトロールの効率化手法を検討する。

2. Instagram における有害投稿の調査

2.1. 予備調査

児童性被害の原因となる有害投稿が Instagram 上にどの程度存在し、その中でどのようなキーワードが用いられているかを確認するため、手動によるハッシュタグ検索を行って検索結果に含まれる有害投稿を調査した。検索には筆者が Twitter 上でのサイバーパトロールを行う過程で把握した用語や隠語を使用した。ただし、Instagram では「援助交際」や「援交」などあからさまな語句のハッシュタグは検索で利用できなかった。これは、事業者が有害と判断した語句による検索を規制しているためと考えられる。

表 1 に 2021 年 11 月 11 日に実行した検索の結果の中で、多数の有害投稿が確認されたハッシュタグの例を示す。

検索結果の投稿をサンプリングして調査した

表 1 多数の有害投稿が確認されたハッシュタグの例

ハッシュタグ	ハッシュタグを含む投稿の総数
p 活	3.4 万
p 活女子	3644
パパ活	3.4 万
パパ活 jk	1816
援募	389
援交垢	300

結果、投稿は以下の 3 種類に分類できることがわかった。

- 有害**：援助交際等の相手を具体的に募集しているもの
- 無害**：子育ての意味で「パパ活」を使用するなど援助交際等とは無関係なもの
- その他**：会員制サイトへの誘導や副業広告など業者によるもの

特に投稿数が突出している「p 活」、「パパ活」のハッシュタグについては、多くの投稿がその他に分類されるものだった。

次に、この予備調査で有害な投稿が確認されたハッシュタグを使用して投稿収集を自動化し、その中の有害投稿の特徴を具体的に調査する。

2.2. 公式 API を使用した投稿の収集

Instagram の公式 API を利用し有害投稿のデータを自動で収集するためのプログラムを Python で作成した。プログラムの概要を図 1 に示す。投稿収集プログラムはキーワードテーブルから検索キーワードを取得し、API を介して検索を行う。そして、検索の結果得られた投稿データを投稿収集テーブルに保存する。



図 1 投稿収集プログラムの概要

2.3. 収集した有害投稿の調査

2021 年 12 月 10～22 日にかけて収集した投稿

の件数と分類結果を図 2 に示す。

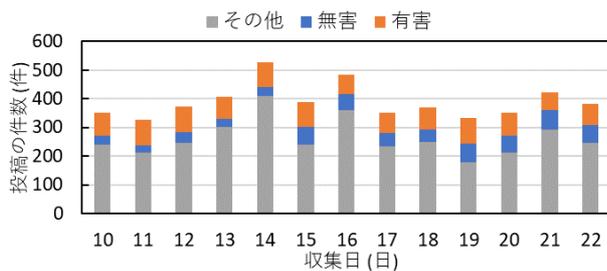


図 2 収集した投稿の分類結果

1 日あたりの平均投稿件数は 390 件、うち有害・無害・その他の投稿件数の平均はそれぞれ 79 件・47 件・264 件であった。よって、実際に有害投稿が多く確認されたため、Instagram でもサイバーパトロールが必要であると言える。また、収集した全投稿のうちその他にあたる投稿は 68%を占める。その他にあたる投稿は児童性被害防止のためのパトロールを行う上では不要な投稿であるため、効率化のためにはこれらを自動で除外する処理も必要になると考えられる。

また、収集した全有害投稿 1027 件の約 46%が、図 3 のような画像内に詳細情報を含む投稿だった。

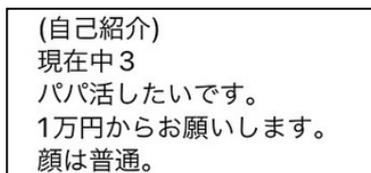


図 3 画像内に詳細情報を含む投稿の例

Twitter を対象とした研究で行われているようなテキスト解析のみでは、このような投稿に十分に対応できないため、効率化の障壁となる。

3. OCR 技術を活用したサイバーパトロール効率化の検討

有害投稿は対象者（援助交際を求めるものであれば、その相手）に意図を伝えるため、隠語などのキーワードを投稿に含む。そのため、Twitter を対象とするテキスト情報を用いた研究が、Instagram にも応用できる可能性が高い。例えば、機械学習で投稿の有害判定を行う研究[2]を応用すれば、効率的に有害投稿を発見できる。ただし、2.3. で示した通り、Instagram では画像内に詳細情報を含む投稿も多く見られた。そこで、画像からテキスト情報を抽出する OCR 技術が活用できると考えられる。

Tesseract という OCR エンジンを用いて、実際に図 3 に示した画像から文字を抽出したところ、図 4 のように正確に抽出できた。

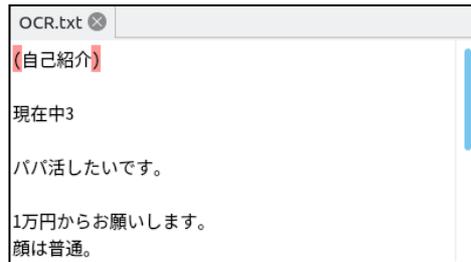


図 4 OCR 技術による文字抽出結果

以上のことから、OCR 技術を活用すれば、画像内に詳細情報を含む場合でも、その情報をテキストデータとして扱えることが確認された。OCR 技術を活用し、Twitter で行われているテキスト情報を用いた研究を応用すれば、Instagram におけるサイバーパトロールの効率化が実現できると考えられる。

4. まとめと今後の課題

本研究では、今後も児童性被害の原因となる有害投稿が増加すると予想される Instagram において、有害投稿がどの程度存在するかを確認し、Instagram におけるサイバーパトロールの必要性を示した。また、有害投稿収集によって得られた知見から、Instagram でサイバーパトロールを行う際の効率化手法について検討し、実際に OCR 技術を用いて画像内のテキストが抽出できることを確認した。今後は、今回検討した手法について、実際のシステムとして実現することが課題となる。

参考文献

- [1] 警察庁，“子供の性被害（児童の性的搾取等）”，https://www.npa.go.jp/policy_area/no_cp/uploads/kodomonoseihigair3.pdf（参照 2021-11-23）
- [2] 安彦ら，“機械学習を用いた薬物売買におけるサイバーパトロールシステムの開発”，情報処理学会論文誌，Vol. 61, No. 3, pp. 535-543 (2020)
- [3] 羽田ら，“コーパス間での類似語の差異に着目したマイクロブログにおける隠語検出”，情報処理学会研究報告，Vol. 2020-ICS-200, No. 2, pp. 1-8 (2020)
- [4] 池辺正典，“サイバー防犯ボランティアのためのサイバーパトロール支援システムについて”，警察学論集，Vol. 71, No. 5, pp. 146-158 (2018)
- [5] 池辺ら，“SNS を対象とした都道府県警察のサイバーパトロール支援システムについて”，情報教育シンポジウム論文集，Vol. 2021, pp. 196-201 (2021)