

オンライン詐欺や犯罪へ誘導する SNS 投稿文の類型化と特徴の分析

松田 美慧^{1,2,a)} 川口 大翔¹ 藤田 彬^{2,3} 吉岡 克成^{3,4}

概要：ソーシャルネットワーキングサービス（以下、SNS）は情報収集やコミュニケーションの手段として人々の生活に浸透したが、見知らぬ人同士が気軽に直接連絡を取り合えることから、意図しない情報の流通やトラブルにユーザを巻き込む危険を伴う。特に短文投稿型の SNS において「高額バイト」などと称して人材を募集し、応募者を特殊詐欺や強盗の実行犯として利用する手口や登録費などとして金銭をだまし取る手口は、身近な脅威として広く認知され、問題視されている。このような手口に対し、ユーザに呼びかけられている主な対策には、犯罪へ誘導する投稿に頻出するキーワード（「高額」「即日現金」「副業」など）に注意することが挙げられるが、この対策では使用されるキーワードが変われば対応が困難となる。本研究では、犯罪やオンライン詐欺に誘導するために発信された可能性の高い投稿を短文投稿型の SNS より収集し、類似する投稿文を類型に纏めたうえで、類型ごとに投稿文の構文上の特徴や推定されるターゲットの人物像、アカウントの繋がり方などの特徴を分析した。その結果、犯罪やオンライン詐欺と関連する可能性が高い投稿文では、文の構成や文言が一致あるいは類似した投稿が多く確認され、複数のアカウント間で投稿をテンプレートのように使いまわしていることが明らかになった。同一の投稿の使用のされ方は引用による再投稿が主であるが、場合によっては人気アイドルグループ名や「子育てママ」など対象者を限定する単語を入れ替えて投稿が繰り返されており、明確なターゲットに届けるように発信内容を工夫していると推測される。

キーワード：ソーシャルメディア、有害情報、テキスト分類、特徴分析

Typification and Characteristics Analysis of Posts Leading to Online Fraud and Crime on Social Media

MISATO MATSUDA^{1,2,a)} YAMATO KAWAGUCHI¹ AKIRA FUJITA^{2,3} KATSUNARI YOSHIOKA^{3,4}

Abstract: Social media have permeated people's lives through information gathering and communication, but their use risks leading users into serious incidents. In particular, the use of social media that post short messages to solicit applicants for "high-paying jobs" and use them as perpetrators of special frauds and robberies, as well as the deception of money as registration fees, are widely recognized as familiar threats and are considered problematic. The main countermeasure against such tactics is to pay attention to keywords that frequently appear in posts leading users to commit crimes (e.g., "high price," "instant cash," "side job," etc.). In this study, we collected posts from short-form social media that are highly likely to be sent out to induce people to commit crimes or online fraud, grouped similar posts into categories, and analyzed the syntactic characteristics of each category, the presumed target persona, and the characteristics of the account connections. In addition, we analyzed the relationship between the number of posts, the number of responses for each type, and the number of accounts that sent the posts. The results showed that many posts that were most likely related to crimes or online fraud had identical or similar sentence structures and wording and that the posts were used as templates among multiple accounts. The same postings were mainly re-posted by citation, but in some cases, postings were repeated by replacing words that limit the target audience, such as the names of popular idol groups or "mothers raising children," suggesting that the content of the postings is designed to reach a clear target audience.

Keywords: Social Media, Harmful Information, Text Classification, Characteristic Analysis

1. はじめに

インターネットの利用の増加に伴い、児童ポルノなどの違法な情報や自殺の誘引などの有害な情報の流通が社会問題となっている。これに対し、2006 年よりインターネット・ホットラインセンターは警察庁の業務委託を受け、一般市民からの違法・有害情報に関する通報を受け付け、都道府県警察が実施する捜査活動あるいはサイト管理事業者への削除依頼を支援している [1]。しかしながら、昨今では Web サイトを入り口とせず、違法・有害情報の対象となっていなかった新たな手口も登場している。なかでも社会に大きなインパクトを与えた事例として、短文投稿型の SNS において「高額バイト」などと称して人材を募集し、応募者を特殊詐欺や強盗の実行犯として利用した手口があげられる。この新たな手口に対し、警察庁は特殊詐欺などの闇バイトを有害情報に加え、人工知能を導入した対策を開始することを発表した [2]。またユーザに向けて、犯罪への加担に誘導する投稿に頻出するキーワード（「高額」「即日現金」「副業」など）に注意を呼びかける対策がなされている。しかしながら、表現やキーワードに着目した対策は簡潔で実行し易い利点があるものの、表現やキーワードを別のものに置き換えられた際に再び投稿文の特徴の分析と対策の練り直しが求められる点で、持続可能性の面で一定の限界があるともいえる。持続可能性を高めるためには表現やキーワード以外に違法・有害情報に共通する特徴を特定する必要があり、投稿文の構文や取り扱われる話題などの上位語、投稿がターゲットとする人物像や投稿アカウントの使われ方などメタ的な特徴などに着目し、ユーザが注意すべき特徴を分析する必要がある。

本研究では、犯罪やオンライン詐欺に誘導するために発信された可能性の高い投稿を短文投稿型の SNS より抽出し、投稿文を類似性の高いものをまとめて類型化した。投稿文の類型に対し、以下の 3 つのリサーチクエスション（以下、RQ）を設定し、分析にあたった。

- RQ1：同じ投稿文を共有するアカウントはあるか
- RQ2：共有するアカウントが複数ある場合は中心となる投稿は存在するか
- RQ3：ターゲットとして推測される人物像や仕事の内容、URL の有無に関連はあるか

投稿文の類型ごとに発信したアカウントのペア群を特定し、アカウントのつながりの集散の具合を可視化した。投稿文の内容の特徴の分析ではターゲットの人物像などを表すラベルを作成し、類型ごとのラベルの有無との相関係数を算出し、関係性を明らかにするように試みた。これらの分析より、ユーザが注意すべき特徴の検討を行った。

2. 関連研究

インターネットで流通する有害情報の検知には、有害情報に指定された正解データから機械学習手法により検知する方法 [3][4] とキーワードリストを利用する方法 [5][6] がある。今回収集したデータは有害情報の可能性が高いが、その悪性を立証する正解データが存在しないため、キーワードリストによる検知の関連研究と悪性の立証を試みる関連研究をそれぞれ取り上げる。池田ら [6] は、キーワードによる違法・有害サイトの検出では、文脈を考慮しないために誤検知が誘発されているとして、キーワードを基にした文節間の係り受け解析を導入したフィルタリングシステムの提案を行った。大規模な Web 文書群を用いた性能評価の結果、池田らの提案手法は従来手法に比べ、判定の誤りを多く検出できたとしている。

川口ら [7] は、短文投稿型の SNS からメッセージ型の SNS へユーザを誘導し、不審なサイトへのアクセスや送金を促すオンライン詐欺の可能性のある活動を定期的に観測するシステムを構築し、短文投稿型およびメッセージ型の SNS から収集した投稿やアカウントなどの情報よりアカウント間の関係を調査した。その結果、短文投稿型の SNS から有害なメッセージ型の SNS の公式アカウントに誘導するアカウント群が特定された。この群は、メッセージ型の SNS から詐欺の疑いのあるサイトへと誘導するのアカウントで、その多くが相互のアカウントにユーザを誘導しあうネットワークを形成していることを明らかにした。

3. 調査手法

3.1 データの収集

本研究では、関連研究 [7] で収集したデータのうち、短文投稿型の SNS の投稿とメッセージ型の SNS のアカウントのデータを用いる。当該関連研究では、短文投稿型の SNS からメッセージ型の SNS、Web サイト、SMS、詐欺の疑いのある Web サイトへ誘導するアカウントを調査し、誘導元と誘導先のペアのアカウント情報を紐づけた。

短文投稿型の SNS の投稿は、同サービスが提供する Web API を利用し、“高額バイト”という検索クエリによって検索して得られた投稿と投稿したアカウントに固定された投稿の情報を収集した。収集期間は 2022/12/15-2023/4/17 で、収集した投稿数は合計 70,013 投稿である。投稿内に URL が含まれる場合は URL を抽出し、その URL がメッセージ型の SNS のアカウントへの招待リンクである場合

¹ 横浜国立大学大学院環境情報学府
Graduate School of Environment and Information Sciences,
Yokohama National University

² 国立研究開発法人情報通信研究機構 National Institute of Information and Communications Technology

³ 横浜国立大学大学院先端科学高等研究院
Institute of Advanced Sciences, Yokohama National University

⁴ 横浜国立大学大学院環境情報研究院
Faculty of Environment and Information Sciences, Yokohama National University

a) matsu-misato-fy@ynu.jp

は、誘導先であるメッセージ型の SNS のアカウント情報を同サービスが提供する Web サイトから収集した。メッセージ型の SNS のアカウント情報の収集期間は 2022/12/21-2023/4/6 の期間で、収集対象のアカウントは合計 210 件であった。関連研究 [7] では、誘導先のメッセージ型の SNS において悪性の高いアカウントの特定を行っており、本研究ではメッセージ型の SNS で悪性があると判断されたアカウントと誘導元である短文投稿型の SNS の投稿を突合したデータを利用する。

3.2 クリーニング処理

データ分析の前処理として、投稿文の類似度の計算に不要な文字の削除と置換を実施した。削除の対象とした不要な文字列は 5 種類あり、短文投稿型の SNS でユーザをメンションする際に表出される「@+アカウント名」で構成される文字列、投稿の文章末にある文脈とは無関係な整数あるいはアルファベットの文字列、他者の投稿を引用して再投稿した場合を示す文字列「RT:」、特殊文字などの記号（例：♪, ★, ☆, !, ? など）と句読点である。なお、アカウント名が文章の冒頭あるいは文章末に位置していない場合に、個別で連絡を取り合う連絡先として記述されているものが確認されたため、「@name」と文字を置換した。その他の不要な文字の置換では 5 種類の文字を置換した。具体的には、大文字を小文字に、全角の英数字および全角の空白を半角文字に、カタカナを全て全角文字に、絵文字をアスタリスク記号へ置換し、URL を「http」と略記した。

3.3 表層による類型化

処理を簡易にするため、投稿文の表層に基づく類型化を行う。類似度の高い投稿文を集約のため、投稿文間でレーベンシュタイン距離を計算し、算出した値が 0.9 以上となる投稿文群を作成した。本来は群内のいずれかの投稿とレーベンシュタイン距離が 0.9 以上になる投稿文群を二次的に参照する作業を繰り返す必要があるが、二次参照の繰り返して類似度の低いものが同一類型に混ざること避けるため、今回は二次参照は行わない。完全に一致する投稿文ではなく類型化を実施した理由は、誤字脱字、絵文字や改行と空白の有無に代表されるような人間が読む際には同一の意味で判断される文章を一つの類型とするためである。なお、今回の方法では完全に一致する語句の出現順序が大幅に異なる場合や文章の構造は類似しているが使用される単語が異なる場合などは同一類型としない。類型への集約には反映されなかった類型間の類似を確認するため、可視化を行う。可視化では、類型を代表する投稿文群の TF-IDF (Term Frequency - Inverse Document Frequency) を算出して文章をベクトル表現に直し、次元削減した結果を 2 次元平面にプロットした。TF-IDF とは文章中に出現する単語の重要度を表し、今回は類型に含まれる単語が他の類型

表 1 収集データの構成

種別	投稿数
1) 人材募集を行う投稿文	59,857 投稿
2) 応募を希望する投稿文	124 投稿
3) 高額バイトへの意見を表明する投稿文	1,898 投稿
4) わいせつな動画を宣伝する投稿文	997 投稿
5) テレビ番組やニュース記事を宣伝する投稿文	3,037 投稿
6) いずれにも分類されない投稿文	4,100 投稿
合計投稿数	70,013 投稿

に含まれない場合に重要度が高くなる。高次元ベクトルの次元削減には t-SNE (t-distributed Stochastic Neighbor Embedding) を利用する。TF-IDF と t-SNE の算出には、Python の既存ライブラリ scikit-learn を用いた。

3.4 分析対象の抽出

違法・有害情報の抽出にはキーワードによるパターンマッチングが一般的であるが、短文投稿型の SNS で違法な仕事やサイトへ誘導するキーワードセットは公表されたものがなく、犯罪との関連を立証できない。しかし、違法・有害情報である可能性の高いものを抽出する必要があるため、本研究では 3 つの独立した方法を試した。その方法は、目視による抽出、オンライン詐欺との関連があるメッセージ型の SNS アカウントとの関連の有無による抽出、犯罪との関連が極めて高いキーワードによる抽出である。

3.4.1 目視による抽出

収集データには、単なるコメントやアルバイトへの応募希望など明らかに何らかの人材募集とは別の意図で発信されたと判断される投稿が混在している。違法・有害情報である可能性の高いものを抽出するため、何らかの人材募集を意図して発信されたと推測される投稿文を抽出した。分析対象を抽出する作業は、原則として人手で実施し、投稿文の内容が何らかのアルバイトの紹介または人材募集であり、かつダイレクトメッセージをはじめとする個別での連絡を最終的に要求するか否かを判定の基準とした。

これらに該当しない投稿文については、その内容に合わせて判別を行い、表 1 のように分類した。本研究では短文投稿型の SNS にて犯罪やオンライン詐欺に関連する可能性の高い投稿を対象とするため、1) の人材募集を行う投稿以外の投稿は分析対象から除外した。勿論、2) ~ 6) に分類される投稿であっても人材募集を目的として発信された可能性はあり、発信した投稿者と連絡を取り合うと間接的に人材募集や参加の呼びかけに至る投稿文が含まれている可能性はあるが、本稿では単一の投稿文だけで直接的に人材募集している投稿に焦点を当てる。

3.4.2 複数 SNS への誘導の有無による抽出

人材募集を行う投稿のうち、実際に違法な活動との関連がある投稿か否かを判断するには短文投稿型の SNS の投稿文だけでは難しい。本研究では投稿の悪性を判断する

根拠の一つとして、関連研究 [7] で分析したメッセージ型の SNS の情報を用いる。関連研究 [7] では、誘導先のメッセージ型 SNS において、相互に誘導を行いつつ様々な副業を勧める「相互誘導のネットワーク」を構成するアカウントグループ (以降では論文 [7] 内の呼称に従い「副業系 1」グループと呼ぶ) を発見した。このグループは不審な決済を促すサイトや個人情報の入力を促すサイトにユーザを誘導し、誘導先の当該サイトに関して被害を報告する口コミが確認されたから、それらが悪意のあるグループである可能性が高いと判断している。そのため、本研究ではそれらのアカウントに誘導する短文投稿型の SNS の投稿も悪意がある可能性が高いと判断し、該当する投稿を抽出した。

3.4.3 犯罪との関連の高いキーワードによる抽出

警視庁が特殊詐欺との関連が高い違法・有害情報として削除対象に指定した表現には「著しく高額な報酬を示唆する表現」および「犯罪の実行者の募集を示唆する表現」がある。収集データは「高額バイト」というキーワードで検索して得られた投稿であるため、既に高額な報酬を示唆することから、その他の条件に該当する投稿か否かを判定する。「犯罪の実行者の募集を示唆する表現」を示すキーワードは、警視庁が違法・有害情報としての上記の文言を追加削除する方針を固めたことについて報じた新聞記事やユーザに対する注意喚起を目的として公的機関が発信した情報より、削除対象あるいはユーザが注意すべきキーワードとして紹介された文言を収集し、これを元に 25 単語で構成されるキーワードリストを作成した。キーワードリストはキーワードを含む種類の件数とともに、付録 A-1 に掲載する。キーワードリストに登録された 2 種類以上の文言が類型に含まれる場合は、該当する投稿として抽出する。含まれる文言が 2 種類以上とすることでできるだけ誤検知を減らすよう工夫しているものの、提案方法はパターンマッチングであり、文脈を考慮しないため、「打ち子の仕事ではありません」などのようにキーワードの内容を否定している場合でもキーワードがあると判定され、逆に「荷受け」と「荷物の受け取り」のようにキーワードと同一の意味を持つがキーワードを完全に含まないものは対象外となる。これらの作業により、関連研究 [7] の方法では網羅できなかった悪性を検証対象とした。

3.5 キーワードリストの作成とラベルの付与

人材募集を行う投稿の特徴を明らかにするため、募集する仕事の内容と募集対象とする人物像に関するキーワードリストを作成し、類型ごとにキーワードの有無を判定した。キーワードリストに含まれるキーワードは、収集した投稿文の文中において改行もしくは空白で区切りられた語句から全ての類型間での重複を削除した 8,882 個の語句を基礎とし、目視にてキーワードを選別し、単語の追加と削除を実施した。キーワードにはあらかじめキーワードの特

表 2 仕事の内容に関するキーワードリストの構成

種類	収容数	例
報酬	166	円、即日即金、時給、日当、月 30、100 万、高単価、昇給あり、寮付き
募集地域	109	東京、場所、海外、どこでも、お家、在宅、自宅、1 都 3 県、一部三県、静大、リゾート
勤務期間	15	単発、短期、長期、一人 1 回まで、期間、月 10 件まで、最短 1 ヶ月、日雇い、派遣
勤務時間	75	夕方から、ナイトワーク、夜勤、育児、自分のペース、9:00-18:00、自由勤務、時間
応募条件	60	意思疎通できる人、稼働条件、開設前もしくは解約出来る方、顔付き身分証、携帯 1 台

徴を表すラベルを付与し、類型にキーワードが含まれる場合には特徴ラベルを類型に付与する。一方で今回の分析では文脈を考慮しないため、第 3.4.3 項のキーワード付与と同様に、誤検知の可能性がある。人物の条件を表す名詞の直後に指定した否定語（不可や NG など）が連続して存在する場合は、あらかじめキーワードリストに名詞と否定語をセットにした語句を登録することで除外処理を行うが、それ以外のものは含まれる。

3.5.1 仕事の内容に関するキーワードリスト

仕事の内容に関するキーワードは、一般的なアルバイトの募集の際に掲載される項目と同様に [仕事内容 / 報酬 / 募集地域 / 勤務期間 / 勤務日数と時間 / 応募条件] の 6 種類に分類される。本研究では上記の 6 種類のうち、仕事内容については第 3.4.3 項にて違法性のある仕事を網羅していることから、仕事内容を除く 5 種類の項目に該当するキーワードを選別し、5 種類のキーワードリストを作成した。バイトを募集している地域に関するキーワードリストは、都道府県名、地方名、政令指定都市名の一覧より重複を排除したものに、投稿文中で抽出された語句の「全国、日本、都内、テレワーク、在宅」などの単語を加え、作成した。それぞれのキーワードリストに登録される語句と語句数を表 3 にて示す。5 種類の項目に該当するキーワードを含むか否かを 2 値で表現する特徴ラベルを類型ごとに付与した。

3.5.2 対象の人物像に関するキーワードリスト

応募対象とする人物像に関するキーワードは、収集した投稿文の文中において改行もしくは空白で区切りられた語句から収集し、人に関連する文言を目視にて抽出した。抽出したキーワードは、[性別 / 年齢 / 金銭を必要とする度合 / 体格 / 精神状態 / 国籍 / 職業] の 7 種類があり、種類ごとにキーワードを収容した 7 つのキーワードリストを作成した。このうち国籍に関するキーワードリストに登録された語句は 3 件と非常に少なかったことから、国籍を除く 6 種類のキーワードリストについて、類型ごとに該当するキーワードの有無を 2 値で表現する特徴ラベルを付与した。一方、性別には [全性別 / 男のみ / 女のみ] の 3 つの場合があり、年齢には [全年齢 / 20 代以下の青年層 / 30 代から 50 代くらいの壮年および中年層] の 3 つの場合がある。そのため、性別と年齢に関してはより詳細な分析を行うため、それぞれの場合に合わせてキーワードを細分類し、類型ごとに細分類したキーワードリスト 3 つに該当するキーワードの有無を 2 値で表現する特徴ラベルを付与した。全性別あるいは全年齢を対象とする場合は、当該キーワード

表 3 対象の人物像に関するキーワードリストの構成

種類	収容数	例
性別_all	20	誰でも、どなたでも、どの様な人でも、ノンケ、カップル、性別関係無し、老若男女
性別_female	90	女、ママ、主婦、JK、h カップ、パパ活、BTS、なにわ男子、整形、担当狂、ホス狂
性別_male	35	男、パパ、サラリーマン、ゲイ、DK、BLACKPINK、m 活、イケメン、パチスロ、メンズ
年齢_all	20	誰でも、どなたでも、年齢不問、18 歳以上、10 代から 40 代まで
年齢_middle	50	会社員、育児、子育て、不妊治療費、ホワイトニング、観劇、高校生不可
年齢_young	95	18 歳 35 歳、若い、高校生、20 代、24 卒と繋がりた、キャバ嬢、BTS、春から岡山
金銭	429	お金、給料が減った方、オンラインカジノ、育児、BTS、ガチャ爆死、稼ごたい、生活変えたい
体格_s	15	h カップ、いい身体、体育会系、筋トレ、ルックス、見た目に自信ある人、スレンダー限定
精神状態	38	メンヘラ、やる気、ヤル気、うつ病、依存症、覚悟、楽しく生きよう、死にたい
職業	77	イベントスタッフ、会社員、社会人、ストックビジネス、せどり、パート、フリーター、医師、職歴不問

リストに該当するキーワードが類型内になく場合、他の 2 つの細分類したキーワードリストによって付与された特徴ラベルが両方ともキーワードが存在することを示す場合は、全性別あるいは全年齢に関する特徴ラベルも該当キーワードがある場合と同等に扱うこととする。以上の作業により、それぞれのキーワードリストに登録される語句と語句数を表 2 にて示す。複数のキーワードリストで重複して登録されるキーワードについては、付録 A.2 に掲載する。

3.6 類型と発信アカウントとの関係の分析

同一の類型を共有するアカウントの数の特定では、同一の類型に属する投稿を発信者のアカウント ID を記憶し、単に重複するアカウント ID を削除した場合と他者の投稿を引用して再投稿したアカウント ID を消去した上で同様の処理をした場合の比較を行った。悪性の高い投稿に対し、同一の類型に複数の異なるアカウントが確認された場合は、python の既存ライブラリ networkx と matplotlib を使用し、アカウント間の繋がりを可視化した。

3.7 悪性の高い投稿の内容の分析

類型ごとに投稿の特徴を表現するラベルを付与し、犯罪とオンライン詐欺との関連が高い投稿に特異的な特徴の有無を分析する。投稿の内容を表現する特徴ラベルは第 3.5 節で説明した仕事の内容と対象の人物像を表す特徴ラベル 12 種類に加え、URL、メンション、ハッシュタグ、再投稿の有無を示す 2 値ラベルの計 16 種類である。この 16 種類のラベルを悪性の高い投稿に付与し、ラベル間の相関係数を求め、相関の有無と強弱により、犯罪とオンライン詐欺との関連が高い投稿文の特徴の分析を行う。

4. 分析結果

4.1 類型の可視化

人材募集を行う全投稿は 59,864 投稿あったが、全文が完全に一致する投稿をまとめると 11,922 種類（全投稿の 19.91%）となり、さらにレーベンシュタイン距離が 0.9 以上の投稿文をまとめると、投稿文は 10,131 種類（全投稿の 16.92%）の類型に集約された。類型同士の関係を t-SNE を用いて可視した結果を図 1 にて示す。図中の色は、青色の三角記号が関連研究 [7] にてオンライン詐欺との関連が高いと判断された「副業系 1」に分類された投稿文を示し、緑色の四角記号は犯罪との関連の高いキーワードリストに

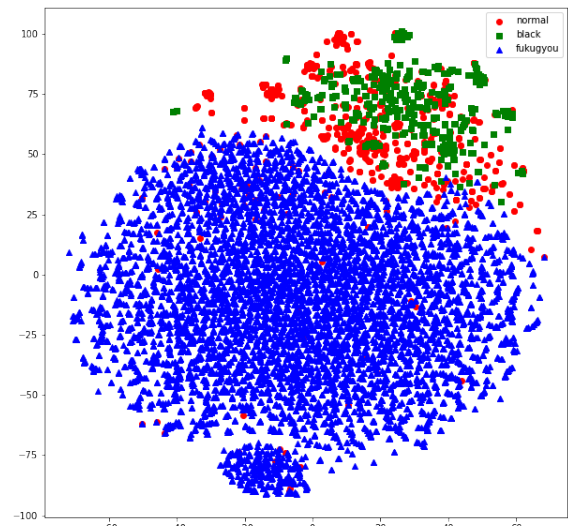


図 1 類型間関係

よって抽出された投稿文を示し、赤色の円記号はそ例外の投稿文を示す。図 1 より、目視によって抽出した人材募集を行う投稿文は複数の小さい類型群が集まった 3 つの大きなグループに分かれることが明らかになった。以降では便宜上の名称として、原点を含む最大の類型群をグループ A、横軸が 0 以上 25 以下かつ縦軸が 50 以上辺りに位置する類型群をグループ B、その他の類型をグループ C とする。オンライン詐欺と関連する可能性の高い投稿は、グループ A および B に偏っており、投稿の類型は 8432 種類あった犯罪と関連する可能性が高い投稿はグループ C に偏って存在し、類型数は種類であった。オンライン詐欺との関連も犯罪との関連も確認されなかった類型は 1699 種類であった。

4.2 同一の類型を共有するアカウントの分析

人材募集を行う投稿を発信しているアカウントの数は 2720 件であった。このうち、他者の投稿を引用して再投稿したアカウントの ID を消去した上で同一の類型内で重複するアカウント ID を削除した場合のアカウント数は全て 1 件となり、同一の類型を共有するアカウントが複数ある場合は他者の投稿を引用して少なくとも一回以上の再投稿をしていることが明らかになった。再投稿を行ったアカウントの数は 2160 件であった。アカウント ID が重複する場合を排除した同一の類型を共有するアカウントの数は、最大 698 件、最小 1 件、平均 1.82 件になった。反対に類型数に注目すると、10131 種類のタイプのうち、複数のアカウントによって再投稿された類型は 451 種類で、その他の 9680 種類は単一のアカウントによって発信されていた。

4.2.1 オンライン詐欺に関連するアカウントの分析

引用を再投稿したアカウントのうち、オンライン詐欺に関連する類型を再投稿しているアカウントの関係を図 2 にて示し、関連のあるアカウント数が多いアカウントの上位 20 位を図 3 にて示す。ネットワーク図より、引用を再投稿

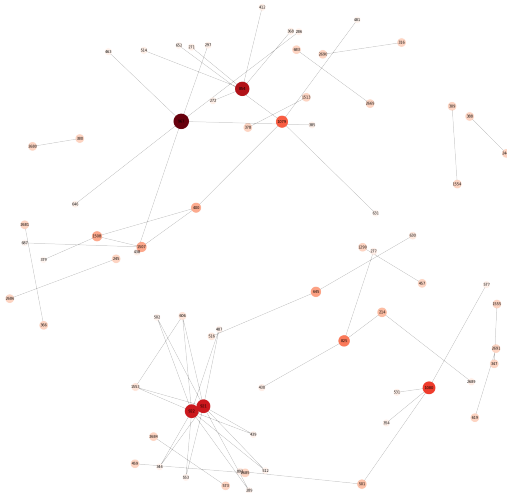


図 2 オンライン詐欺と関連のある類型を共有するアカウント間の関係図

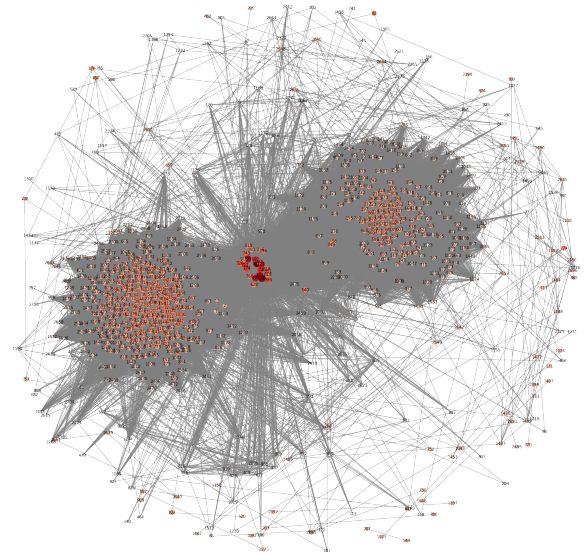


図 4 犯罪と関連の高い類型を共有するアカウント間の関係図

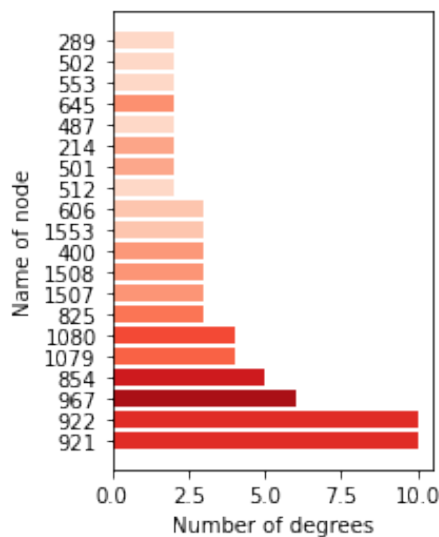


図 3 オンライン詐欺と関連のある類型を共有するアカウントの上位 20

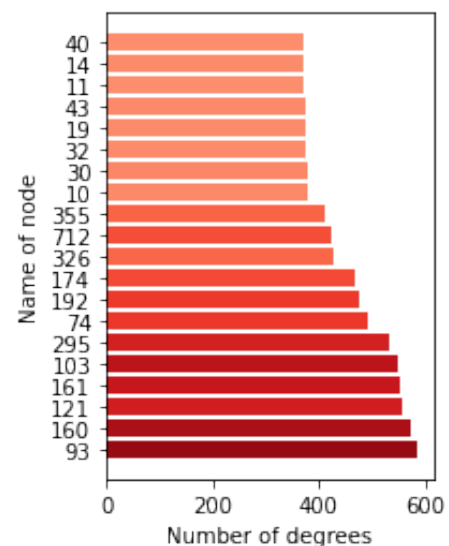


図 5 犯罪と関連の高い類型を共有するアカウントの上位 20

したアカウントは最大 10 件のアカウントと投稿を共有していたが、2つのアカウント間でしか共有をしていないペアが 10 組あり、オンライン詐欺との関連が疑われるアカウント全てが繋がりを持ってはいないことが明らかになった。

4.2.2 犯罪に関連する可能性の高いアカウントの分析

引用を再投稿したアカウント同士の関係を図??にて示し、関連のあるアカウント数が多いアカウントの上位 20 位を図??にて示す。ネットワーク図より、引用を再投稿したアカウントは少なくとも 3つの投稿を共有するグループがあり、1つのグループがその他 2つのグループをつなぐ中心的な役割を果たしていることが明らかとなった。

4.3 投稿の内容の分析

犯罪とオンライン詐欺との関連の高い投稿の特徴を明らかにするため、各特徴ラベルとの相関係数を算出した結果を表 4 にて示す。表中においてオンライン詐欺と関連する可能性が高い投稿群は [副業系] の列、犯罪との関連する可能性が高い投稿群は [犯罪]、上記のいずれにも該当しない投稿群は [その他] の列に結果を示す。投稿の内容を表すラベルは、1) 募集する仕事内容に関する特徴ラベルの [reward / locate / term / time / require]、2) 募集する対象の人物像に関する特徴ラベルの [gender(all,both,female,male) / age(all,both,middle,young) / money / physique / mind / status]、3) クリーニング処理で除外した特徴 (ハッシュ

表 4 特徴ラベルとの相関関係

ラベルの種類	副業系	犯罪	その他
hashtag	-0.26	0.15	0.19
mention	0.59	-0.57	-0.29
https	-0.11	0	0.12
RT	-0.11	0.07	0.08
reward	-0.13	0.16	0.05
locate	-0.32	0.27	0.19
term	-0.05	0.12	-0.01
time	-0.03	-0.01	0.04
require	-0.1	0.02	0.09
gender_all	0.05	-0.06	-0.01
gender_both	-0.02	0.01	0.02
gender_female	0.01	0	-0.01
gender_male	0.01	-0.02	0
age_all	0.1	-0.13	-0.03
age_both	-0.03	0.02	0.03
age_middle	0.1	-0.12	-0.03
age_young	0.02	-0.06	0.01
money	0.11	-0.03	-0.11
physique	-0.1	0.05	0.08
mind	-0.04	0.08	-0.01
status	0.03	-0.02	-0.02

タグ、メンションするアカウント、URL、再投稿)の有無を示す特徴ラベル [hashtag / mention / https / RT] である。

相関係数を比較すると、副業系に属する投稿文は [hashtag / mention] との間で正の相関、[locate] との間で負の相関がみられ、犯罪に属する投稿文は [locate] との間で正の相関、[hashtag] との間で負の相関がみられた。一方、その他の募集する仕事の内容と募集する対象についての特徴ラベルとの間には、いずれも相関関係が見られなかった。

5. 考察

5.1 同じ投稿文を共有するアカウントの実態

オンライン詐欺と犯罪に高い関連のある投稿群において類型を共有するアカウントを分析した結果、複数のアカウントが同一の投稿を発信していることが明らかになった。共有の仕方は他者の投稿を引用し再投稿によるもので、投稿数の多い類型は発信したアカウント自身が自分の投稿を再投稿する場合も多く確認された。同じ投稿を共有しあう複数のアカウントは、オンライン詐欺に関連する投稿文群では中心的な役割を果たすアカウントは確認できなかったが、犯罪に関連する投稿文群では中心的役割をするアカウント群が発見された。これらのことから、キーワードで抽出した投稿群のアカウントの繋がりには組織的な強い繋がりがあ、犯罪との関連が立証されなくても同類型を共有するアカウントに着目をすれば、一括して有害情報を共有するネットワークを検知できる可能性があると考えられる。

5.2 ユーザが投稿を閲覧する際に注目すべき特徴量

短文型の SNS で人材募集を行っている投稿は、表層表現が類似している投稿をまとめると全投稿の 16.92% まで類型の数を集約できたことから、同一のテンプレートをほぼ加工せず拡散している可能性が高い。一方で類型の中には、投稿文は部分的に一致しているものの、文末や文頭のハッシュタグを置換、および一部の文章を置換あるいは削除している場合が多くみられた。特にオンライン詐欺との関連の高い投稿については、複数の投稿文で部分一致するフレーズが多く確認された。全体の構成においても類似点の確認され、1 文目でターゲットを限定し、2 文目以降でアルバイトの募集を呼びかけ、文末には矢印記号もしくは三角記号の直後に 2 個から 4 個の単語を列挙する形式がみられた。文末に列挙された単語は、多くの金銭を要するであろう事象に関連するものが多く、芸能人のグループや美容関連、ギャンブルや生活困窮の関連など具体的かつ対象者を限定していた。相関関係は確認できなかったものの、発信者は単にテンプレートを使いまわすのではなく、文末の単語を中心にターゲットに合わせた投稿文の加工が行われている可能性が高いと考察される。また地域名の包含の有無を示すラベルと負の相関関係が示されたことから、オンライン詐欺との関連の高い投稿は募集対象の地名に対する記述は少なく、オンラインでの活動を象徴するような特徴が見られたことから、アルバイトを募集する具体的な地名の有無に注目することが有効な対策となる可能性がある。

6. まとめ

本研究では、犯罪やオンライン詐欺に誘導する可能性の高い投稿を短文型の SNS より収集および抽出し、類型化を行った。類型に基づき、文章とアカウントのが持つ特徴を分析した結果、犯罪やオンライン詐欺に関連する可能性の高い投稿群では他者の投稿を引用し再投稿をするアカウントのコミュニティがあり、投稿の言い回しを変化させて投稿が繰り返されていることが明らかとなった。言い回しの変更の規則性には、文全体の削除や順番の入れ替えが確認されたとともに、人気アイドルグループ名や「子育てママ」など対象者を限定する単語を入れ替えて発信する場合がオンライン詐欺に誘導する可能性の高い投稿群において、特に多く多く確認された。このことから発信者は年齢や性別といった明確なターゲットを定め、ターゲットに合わせてハッシュタグや関連するキーワードを少しずつ変えて投稿されていると推測される。またユーザが注意すべき特徴には地名があることが新たに発見された。短文投稿型の SNS における犯罪の実行役の募集およびオンライン詐欺への誘導は、その有害性を立証する証拠が極めて少ないこと、新たなキーワードへの対応に課題が残る。

謝辞 本研究の一部は JSPS 科研費 22H03588 の助成を受けて行われた。

参考文献

[1] 警察庁, インターネット上の違法情報・有害情報対策, 入手先
{<https://www.npa.go.jp/bureau/cyber/countermeasures/illegal-info.html>} (2023.08.09).

[2] 日本経済新聞, 闇バイト募集情報、削除依頼の対象に追加へ 警察庁, 入手先
{<https://www.nikkei.com/article/DGXZQOUE016420R00C23A8000000/>} (2023.08.10).

[3] Guermazi,R., Hammami,M. and Hamadou,A. : Combining Classifiers for Web Violent Content Detection and Filtering, Proc. 7th International Conference on Computational Science, pp.773-780, Springer-Varlag(2007).

[4] 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, 情報処理学会論文誌 Vol.54 No.2 pp571-584(2013).

[5] 伊藤圭吾, 荒澤孔明, 服部峻: 誹謗中傷による被害を減らすためのツイートにおけるトゲワード検出, 電子情報通信学会 信学技報 (2021).

[6] 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, 第2回データ工学と情報マネジメントに関するフォーラム (2010).

[7] 川口大翔, 高田一樹, インミンパパ, 田辺瑠偉, 吉岡克成, 松本勉: 複数の SNS を悪用したオンライン詐欺の観測に向けた検討, 電子情報通信学会 信学技報 (2023)

付 録

表 A.1 犯罪に関連する可能性の高いキーワードリスト

キーワード	件数
受け子	2953
出し子	2513
打ち子	2644
買い子	133
受け出し子	71
引き子	67
売り子	55
かけ子	3
掛け子	5
並びバイト	2
並び子	42
荷受け	3512
渡すだけ	1
運ぶだけ	26
手押し	83
アカウント買取	81
叩き	779
運び	1412
リスクあり	142
黒のお仕事	2918
窃盗	16
運び屋	46
簡単な仕事	348
闇バイト	43201
裏バイト	38846

表 A.2 対象の人物像に関するキーワードリストの例

キーワード	性別_all	性別_female	性別_male	年齢_all	年齢_middle	年齢_young	金銭	体格	精神	職業
カップル	1	0	0	0	0	1	0	0	0	0
童でも	1	0	0	1	0	0	0	0	0	0
老若男女	1	0	0	1	0	0	0	0	0	0
育児	1	0	0	0	1	0	1	0	0	0
子育て	1	0	0	0	1	0	1	0	0	0
性別も年齢も関係ありません	1	0	0	1	0	0	0	0	0	0
応募条件一切なし	1	0	0	1	0	0	0	0	0	1
職歴年齢不問	1	0	0	1	0	0	0	0	0	1
フリーター	0	0	0	0	0	0	1	0	0	1
不妊治療費	0	0	0	0	1	0	1	0	0	0
黒ずみ	0	1	0	0	1	0	1	0	0	0
p 活	0	1	0	0	0	1	0	0	0	0
キャバ嬢	0	1	0	0	0	1	0	0	0	1
風俗嬢	0	1	0	0	0	0	0	0	0	1
BTS	0	1	0	0	0	1	1	0	0	0
h カップ	0	1	0	0	0	0	0	1	0	0
TravisJapan	0	1	0	0	0	1	1	0	0	0
エステティシャン	0	1	0	0	1	0	1	0	0	0
お店に不満がある姫	0	1	0	0	0	1	1	0	1	1
キンプリ	0	1	0	0	0	1	1	0	0	0
ジャニーズタ	0	1	0	0	0	1	1	0	0	0
主婦	0	1	0	0	1	0	0	0	0	1
シングルマザー	0	1	0	0	1	0	0	0	0	0
ダイエツト	0	1	0	0	1	0	1	0	0	0
なにわ男子	0	1	0	0	0	1	1	0	0	0
ネイル	0	1	0	0	1	0	1	0	0	0
バーノン	0	1	0	0	0	1	1	0	0	0
パパ活	0	1	0	0	0	1	0	0	0	0
ヘアメイク	0	1	0	0	1	0	1	0	0	0
ホス狂	0	1	0	0	0	1	1	0	0	0
ホワイトニング	0	1	0	0	1	0	1	0	0	0
マツエク	0	1	0	0	1	0	1	0	0	0
メルレ	0	1	0	0	0	1	1	0	0	0
整形	0	1	0	0	1	0	1	0	0	0
担当狂	0	1	0	0	0	1	1	0	0	0
ホスホス	0	0	1	0	0	1	0	0	0	0
サラリーマン	0	0	1	0	1	0	0	0	0	0
パチスロ	0	0	1	0	1	0	1	0	0	0
パチンコ	0	0	1	0	1	0	1	0	0	0
パパ	0	0	1	0	1	0	0	0	0	0
ママ	0	1	0	0	1	0	0	0	0	0
ママ活	0	0	1	0	0	1	0	0	0	0
競艇	0	0	1	0	1	0	1	0	0	0
競馬	0	0	1	0	1	0	1	0	0	0
競賞	0	0	1	0	1	0	1	0	0	0
三連単	0	0	1	0	1	0	1	0	0	0
賞金ギリラ	0	0	1	0	1	0	1	0	0	0
賞金ルーム	0	0	1	0	1	0	1	0	0	0
新台	0	0	1	0	1	0	1	0	0	0
宝くじ	0	0	1	0	1	0	1	0	0	0
万馬券	0	0	1	0	1	0	1	0	0	0
16 歳	0	0	0	0	0	1	1	0	0	0
18 歳以上	0	0	0	1	0	0	1	0	0	0
依存症	0	0	0	0	0	0	1	0	1	0
見た目に自信ある人	0	0	0	0	0	0	0	1	1	1
体育会	0	0	0	0	0	0	0	1	1	0
弁護士	0	0	0	0	0	0	1	0	0	1
働きたくない	0	1	0	0	0	0	1	0	0	0
満足していない	0	0	0	0	0	0	1	0	1	0