

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

安彦 智史^{1,a)} 加藤 諒² 北川 悦司³

受付日 2019年5月30日, 採録日 2019年11月29日

概要: 近年のインターネット社会では, SNS (Social Networking Service) の利用による薬物売買や児童ポルノ, 自殺補助といった様々なサイバー犯罪が問題視されている. 特に薬物売買は, インターネットの匿名性を利用した売買手法について潜在化や巧妙化が進行しており, 平成 29 年では薬物利用による検挙人員の約半数は青少年であった. そのため, 厚生労働省では, 警察庁と連携し, 有害情報をフィルタリングするシステムの導入を促進しているが, 検挙人員の低下にはつながっていない. さらに公開型の SNS を中心に, 薬物売買を示唆するアカウントが頻繁に作成されているため, 個別に対応するだけでは根本的な解決につなげることは難しいという課題がある. そこで, 本研究では, 特に重要度が高いとされるマイクロブログを対象に, 機械学習を用いて薬物売買を行うユーザの自動抽出を試みる. さらに, ユーザ間の関係から, 薬物売買に興味があるコミュニティを包括的に把握可能なシステムを開発する. 実証実験の結果, 薬物売買に関連するユーザを高精度に抽出し, ユーザの有害判定やネットワークの可視化を実現できることを証明した.

キーワード: SNS, マイクロブログ, サイバーパトロール, 機械学習, 薬物

Development of Cyber Patrol System in Drug Dealing Using Machine Learning Model

SATOSHI ABIKO^{1,a)} RYO KATO² ETSUJI KITAGAWA³

Received: May 30, 2019, Accepted: November 29, 2019

Abstract: With the spread of SNS (Social Networking Service) in recent years, cybercrime such as drug dealing, child pornography and aiding and abetting a suicide are regarded as a problem. Particularly in drug dealing, about half of the arrested persons in 2017 are adolescents, due to the anonymity of the Internet. Therefore, the Ministry of Health, Labor and Welfare is promoting the introduction of a system that filters harmful information in cooperation with the National Police Agency. However, the number of arrests has not decreased. Therefore, we built a method to automatically detect drug users using a machine learning model for microblogging that seems to be particularly important. Furthermore, we developed a system that can detect the community interested in drug dealing from the relationship between the detected users.

Keywords: SNS, microblog, cyber patrol, machine learning, drug

¹ 仁愛大学人間学部
Faculty of Human Studies, Jin-ai University, Echizen, Fukui 915–8586, Japan

² 京都橘大学現代ビジネス学部
Faculty of Contemporary Business, Kyoto Tachibana University, Kyoto 607–8175, Japan

³ 阪南大学経営情報学部
Faculty of Management Information, Hannan University, Matsubara, Osaka 580–8502, Japan

a) abiko@jindai.ac.jp

1. はじめに

近年, スマートフォンを利用した SNS (Social Networking Service) での被害が増加傾向にある. 警察庁が発表した「平成 29 年における SNS 等に起因する被害児童の現状と対策について [1]」によると, Twitter に起因する被害児童数は, 平成 28 年中期だと 446 人に対し, 平成 29 年中期

になると 695 人と約 1.5 倍に増加している。また、学生専用コミュニティの「ひま部 [2]」でも同様に 77 人から 181 人と被害児童数が増加しており早急な対策が求められている。このような SNS を起点とした事件では、児童買春や児童ポルノ、強制猥褻、強姦、自殺幫助、薬物売買など未成年に限らず様々な年代の利用者が事件に巻き込まれている。この中でも、薬物売買は、インターネットの匿名性を利用した売買手法について潜在化や巧妙化が進行している。厚生労働省が公表した「第五次薬物乱用防止五か年戦略 [3]」によると、平成 25 年時の検挙人員は 1,616 人であったが、平成 29 年には過去最多となる 3,218 人が検挙され、その約半数が青少年であった。この原因として、SNS 上では、薬物に有害性がないといった誤情報が氾濫しており、青少年の大麻乱用の拡大につながっているとされている。この対策として、厚生労働省では、警察庁と連携し、薬物に関する有害情報のフィルタリングの導入を促進している。しかし、比較的フィルタリングの導入率が高いと考えられる青少年においても導入率は過半数に届いておらず [4]、十分な導入状況とはいえない。さらに、薬物売買に関する情報が公開型の SNS に頻繁にあがっていることで、薬物に興味のあるユーザと薬物の売り手とが容易にコンタクトをとれる環境が整っていることも憂慮すべき事項である。これら薬物売買に関する書き込みは、SNS の規約違反や法令違反に該当する可能性があるため、管理者に通報することで記事削除や投稿アカウントの凍結を行うことが可能である。しかし、隠語による書き込みにより通報を逃れるケースや、アカウントが凍結されても、新しいアカウントが作成されるため、書き込みやアカウントを個別に対応するだけでは根本的な解決につなげることは難しい。

これらの社会的課題解決に向けて、本研究では、近年薬物売買による検挙人員数が特に増加しているマイクロブログを対象に、有害ユーザの把握とそのつながりを可視化する仕組みを提案する。隠語を含む薬物売買に関連する書き込みを学習し、有害判定モデルを構築することで、ユーザごとの有害度を算出する。さらに、有害ユーザのつながりを可視化することで、有害ユーザをコミュニティ単位で把握可能なサイバートロールシステムの開発を目指す。

2. 既存研究

著者らは、青少年の安心安全なネット利用実現を目的として、サイバートロールシステムに関する研究開発 [5], [6], [7], [8], [9], [10] を進め、社会実装してきた。そのほかにも、サイバートロールにおける有害情報の抽出に関する研究では、キーワードの出現頻度に基づき判定する手法 [11] や機械学習を用いた判定手法 [11], [12], [13], [14], [15], [16], [17] が提案されている。キーワードの出現頻度に基づき判定する手法 [11] では、有害情報に関するキーワードを辞書として登録し、その一致

件数に基づき SNS などの Web ページが有害であるかを判定する。しかし、インターネット上には、有害情報に関するキーワードが膨大に溢れており、すべてを解析するには多大な処理時間が必要になる。そのため、この研究 [11] では、キーワードを用いたフィルタリングだけでなく、機械学習を用いた有害情報の抽出手法もあわせて提案している。

近年では、機械学習を用いて、自動的に有害情報を抽出する手法 [11], [12], [13], [14], [15], [16], [17] が主流である。これらの手法では、教師データを解析して有害情報に関する特徴を学習し、その特徴に基づき未知のデータが有害であるかを判定している。利用する特徴は研究によって異なり、キーワードを用いる手法 [11], [12], [13], [14] やキーワードの組合せを用いる手法 [15], [16], Web ページの HTML タグ構成の特徴を用いる手法 [17] などがある。機械学習には、Random Forests [18] や SVM (Support Vector Machine) [19], Naïve Bayes Classifier などの分類器を利用している。これらの研究を薬物売買の投稿特徴にあわせてサイバートロールに利用することで、監視対象の大量の投稿記事から自動的に有害情報が含まれる情報を抽出できる。

本研究では、キーワードの出現頻度の特徴を学習し、有害判定する手法を検討する。薬物売買に関する投稿は、SNS の普及により近年急速に増加したものであり、薬物売買の投稿を学習し、有害判定を行っている研究は確認できなかった。そこで本研究では、まず、薬物売買に関する投稿記事を教師データとして、複数の分類器を試すことで、精度を確認する。そして、精度が高かった分類器を用いて、薬物売買を行うユーザの抽出とそのユーザ間のつながりを可視化できるか明らかにする。

3. 研究の概要

3.1 研究システムの概要

本システムの処理フローを図 1 に示す。入力は、マイクロブログの起点ユーザ、出力は、有害度を視覚的に確認可能なユーザネットワークである。本システムは学習モデル構築機能とユーザ情報収集機能、有害度算出機能で構成される。学習モデル構築機能では、マイクロブログ上で発信されている薬物に関する投稿と無害な投稿をキーワードの出現頻度を用いて学習する。ユーザ情報収集機能では、入力データである起点ユーザのフォロー情報を収集し、各フォローの投稿を収集する。そして、有害度を算出するユーザを設定する。有害度算出機能では、判定対象ユーザの投稿群を解析し、構築した学習モデルを参照することで、そのユーザの有害度を算出する。最後に、判定したユーザ群のデータをデータベースに格納し、その情報からユーザネットワークに可視化する。本システムを開発することにより、SNS 上における膨大な投稿から薬物に関するやりとりを自動的に抽出可能になる。さらに、有害ユーザのコ

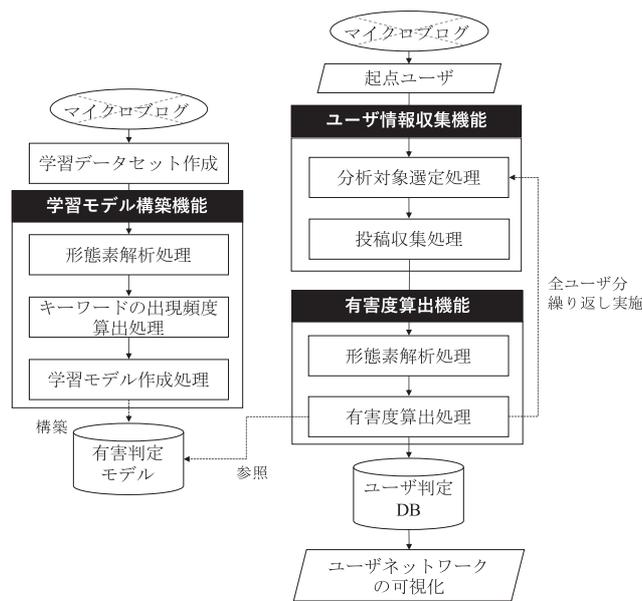


図 1 処理の流れ

Fig. 1 Flow of process.

コミュニティを視覚的に確認できることから、捜査やサイバーパトロールの効率化に寄与できると考える。

なお、本研究では、マイクログログの中でも薬物に関する投稿数の多い Twitter を対象とする。

3.2 薬物売買に関する投稿の特徴

本研究では、薬物売買に関する投稿の特徴を明らかにするため、Twitter から薬物に関する投稿をランダムに 1,000 件収集し、その特徴を分析した。

●フォロー/フォローワ関係の重要性

薬物売買は、ブローカのユーザが薬物の購入を希望する特定のユーザのみを対象 (1 対 1 or 少) とするため、限られたコミュニティに対してのみ情報を発信している。既存研究 [20] では、Twitter 上でヘイトスピーチを行う有害ユーザの検出方法として、リツイートの重要性を述べている。この理由として、ネットいじめやヘイトスピーチを行うユーザは、その性質上、大多数のユーザに投稿内容を拡散 (1 対多) することを目的としている。そのため、ヘイトに共感したリツイートを行ったりするユーザ関係が重要だと考えられる。しかし、薬物売買を対象とした場合、ネットいじめやヘイトスピーチをリツイートするユーザとは目的が異なり、拡散することに意味をなさない。そのため、薬物売買に関する投稿は、ネットいじめやヘイトスピーチに関する投稿と特徴が異なり、リツイート数は少なく、解析対象に適さないことが分かった。

●アカウントの凍結・削除と再作成

ブローカのユーザが使用するアカウントは、有害情報の投稿により Twitter 社から凍結されることや、ユーザの特定を回避するため、ユーザ自身で新規アカウントに変更することが多く、切替えが非常に速いという特徴がある。

実際に、1,000 件の投稿を収集した 10 日後には、284 件の投稿がアカウントの凍結・削除により閲覧できない状況になっていた。これが原因で、アカウントを特定できても、検挙までつなげられないケースが多い。

収集した投稿を分析すると、薬物の買い手側は、薬物購入にのみ使用するアカウントを作成し、複数のブローカとフォロー/フォローワの関係になっていることが分かった。そのため、起点となる有害ユーザのフォローワ関係を対象にすることで、薬物売買のネットワークを可視化でき、検挙につながるコミュニティを検出できると考える。これは、既存研究 [20] でも有用であることが示されており、薬物売買を扱う本研究では特に相関が高いと考えられるため採用した。

●利用頻度の高い隠語と低い隠語が存在

薬物売買では、ヘイトスピーチやネットいじめを対象とした既存研究 [20], [21] と異なり、特定範囲内でのみ通用し、一般とはかけ離れた意味を持たされた言葉である隠語を用いた投稿が非常に多い。ブローカがこの隠語や薬物の商品名を文章やハッシュタグに記載して投稿することで、買い手側のユーザは、ハッシュタグ検索によりブローカを発見し、フォロー/フォローワの関係を構築している。実際に収集した投稿を分析すると、隠語や薬物の商品名を含む投稿は、1,000 件中 981 件と非常に多い。これらの投稿では、全国の薬物売買関係者に認識されている利用頻度の高い隠語をハッシュタグなどに記載して、薬物の買い手側がブローカを発見する手段としている。そして、商品名や取引場所、その他の内容を示す利用頻度が低い隠語をあわせて本文に記載することで、どこで何をいくらで取り扱っているかを示している。記載された条件に合った買い手側のユーザは、DM や匿名性の高い通信アプリを通じて具体的なやりとりにつなげている。このような隠語のみのやりとりにより、不特定多数に情報公開を行い、その後の詳細な売買については、匿名性が高いクローズドな SNS で行っている。このような特徴は、ID 交換掲示板 [8], [9] の特性に酷似していることが分かった。

3.3 学習データセット作成

これらの特徴をふまえ、薬物売買ユーザの検出を目的として、機械学習を行うための教師データを作成する。本研究はマイクログログの投稿内容を対象としていることから、学習データセットは、マイクログログから投稿内容を収集し、薬物売買を示唆する投稿群とその他の無害な投稿群の 2 種類のデータセットを作成する。

3.4 学習モデル構築機能

学習モデル構築機能では、本サイバーパトロールシステムにおいて、ユーザの有害度を判定するための学習モデルを構築する。以下にシステムの詳細を記載する。

3.4.1 形態素解析処理

形態素解析処理では、入力データとなるマイクロブログより収集した学習データセットの投稿群を対象として、MeCab [22] を用いた形態素解析を行い、単語に分割する。なお、単語のうち、投稿内容に含まれる顔文字や平仮名、片仮名1文字といったノイズを取り除くため、形態素が名詞のものを採用した。既存研究 [23] では、電子掲示板を対象に様々な記事のリンクをたどり、有害情報を検出している。電子掲示板を対象としたこの研究では、マイクロブログの特徴と異なり、1件の投稿が長文になる場合も多いため、名詞以外にも形容詞や形容動詞などのすべての品詞を学習データに用い、TF-IDFの手法を基に単語の重要度を算出している。しかし、マイクロブログ上で薬物売買を行うユーザの検出を目的とした場合、すべての記載が短文であるという特徴のほか、ハッシュタグなどに用いられる隠語や取引場所、薬物の商品名など、重要となる品詞の多くが名詞であること、さらにその出現頻度が最も重要であると判断した。

3.4.2 キーワードの出現頻度算出処理

3.4.1項に記載した理由により、キーワードの出現頻度算出処理では、形態素解析により抽出した名詞群に対してBoW (Bag of Words) の手法を用いて単語ごとのラベルを割り振り、特徴ベクトルを作成する。そして、単語の出現頻度から特徴ベクトルを作成する。また、3.2節の事前調査の結果から、利用頻度の高い隠語については、県警と連携して薬物売買によく利用される単語を調査・分析し、そのキーワードを事前に形態素解析辞書に登録した。

3.4.3 学習モデル作成処理

学習モデル作成処理では、キーワードの出現頻度算出処理により作成した特徴ベクトルを用いて分類器による学習を行い、有害判定モデルを構築する。本研究で利用する分類器については、4.1節の事前実験により選定する。

3.5 ユーザ情報収集機能

ユーザ情報収集機能では、分析対象選定処理と投稿収集処理を行うことで、有害度算出機能で判定対象となるユーザとその投稿記事を順次取得する。

3.5.1 分析対象選定処理

分析対象選定処理では、まず、入力された起点ユーザのフォローとして登録されているユーザを収集する。まず、起点となるマイクロブログのユーザIDを入力とし、起点ユーザのフォローとなっているユーザID群を取得する。そして、鍵付きアカウントや投稿をしていないユーザIDを除き、分析対象リストを作成する。

3.5.2 投稿収集処理

投稿収集処理では、分析対象リストに登録されたユーザの投稿内容を収集する。なお、本研究ではシステムの処理負荷軽減のため、取得する投稿数は各ユーザの最新投稿

200件とし、フォロー数が1,000人を超えるユーザは対象外とした。ユーザIDと投稿の収集には、TwitterAPI [24] を用い、マイクロブログのサービスに大きな負荷を与えないシステムの運用を行った。

3.6 有害度算出機能

有害度算出機能では、ユーザ情報収集機能により収集したユーザごとの投稿内容に対して、形態素解析処理、ラベル付与処理を行い、有害判定モデルを用いた二値分類を行う。そして、ユーザごとの投稿総数と有害判定された投稿数の割合をユーザの有害度とし、判定結果をユーザ判定データベースに格納する。

4. 実証実験

実証実験では、まず、事前実験として、本研究の有害判定モデルを構築するための分類器の選定を行う。次に、構築した有害判定モデルを用いた精度実験を行う。最後に、起点ユーザを入力として、薬物売買に関するユーザネットワークの可視化を行う。

4.1 事前実験

4.1.1 実験内容

事前実験では、投稿の有害判定において利用する分類器ごとの精度を確認する。分類器には、従来研究 [23] でネットいじめの自動検出のために用いられたRandom Forests, SVMとNaïve Bayes Classifierの精度比較を実施する。SVMのパラメータであるカーネルには、文書分類で一般的に用いられる線形カーネルを設定した。また、Naïve Bayes Classifierでは、2クラス分類のGaussian Naïve Bayesを採用した。事前実験で使用するデータセットには、サイバー防犯ボランティアにより収集されたマイクロブログの薬物に関する投稿2,094件とその他の無害な投稿5,870件の計7,964件を用いる。これらの正解データは、正確な学習モデルを構築するため、人手により分類して作成した。この実験データを対象に10分割交差法を用いて、学習データと判定データを交差させ、投稿ごとの有害判定の精度を確認する。なお、事前実験では、判定データの正解であった割合をスコアとして算出した。

4.1.2 実験結果

実験結果を表1に示す。表1に示すとおり、機械学習において、分類器にRandom Forestsを利用した場合の精度が最も高く、次いでSVM, Naïve Bayes Classifierとなった。Random Forestsは集団アルゴリズムであり、ランダムサンプリングされた訓練データから学習した多数の決定木を使用する手法である。本研究で扱うようなノイズの多いデータに対して僅差ではあるがSVMよりも高い性能を示したため、本研究では、Random Forestsを分類器としたシステムを構築し、次節以降の実験でも採用する。

表 1 学習器ごとの 10 分割交差法のスコア結果

Table 1 Result of score of 10 Cross-validation each learning machines.

	1回目	2回目	3回目	4回目	5回目	6回目	7回目	8回目	9回目	10回目	平均
Random Forests	0.943	0.934	0.920	0.937	0.933	0.937	0.944	0.928	0.928	0.927	0.933
SVM	0.946	0.929	0.929	0.942	0.929	0.933	0.934	0.920	0.919	0.912	0.929
Naïve Bayes Classifier	0.856	0.878	0.858	0.854	0.837	0.865	0.880	0.851	0.845	0.845	0.857

表 2 有害情報抽出の結果

Table 2 Result of Harmful information detection.

適合率	再現率	F 値
0.991	0.785	0.876

4.2 有害度算出の精度実験

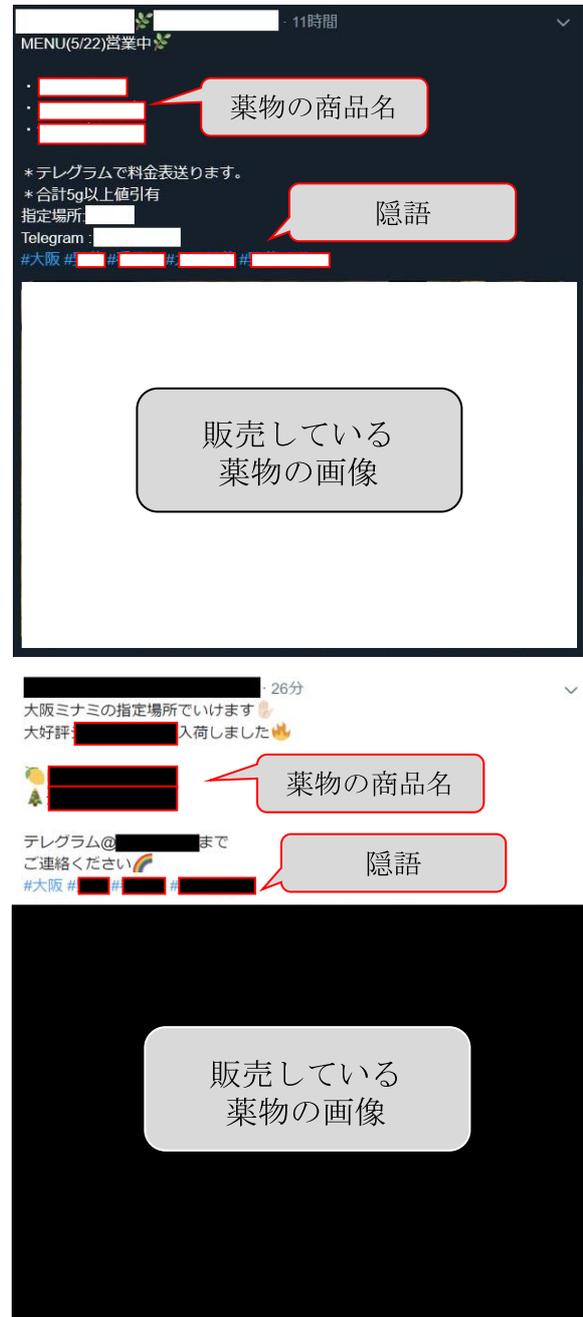
4.2.1 実験内容

本実験では、構築した学習モデルを用いて、ユーザの投稿ごとの薬物売買に関する有害判定の精度を検証する。有害判定に使用するデータは、事前実験に用いたデータとは異なる有害ユーザの 887 件の投稿を用いる。これらの投稿を対象に学習モデルを参照して、有害・無害の自動判定を行う。

判定する際の有害な投稿の基準としては、投稿内に薬物を示す単語や薬物の金額や販売を示唆している単語が含まれている場合とした。本実験の精度は、適合率、再現率と F 値を用いて評価する。

4.2.2 実験結果

実験結果を表 2 に示す。その結果、適合率 0.991 ポイント、再現率 0.785 ポイントと F 値 0.876 ポイントとなり、高精度な分類が可能であることを確認できた。適合率の結果から、有害な投稿は、非常に高い精度で抽出できることが分かった。しかし、再現率は、適合率と比較して低い結果となった。この原因としては、薬物売買に関する詳細情報をテキストで記載せず、画像に載せて投稿しているケースや絵文字と数値のみを記載しており、プロフィールに詳細を投稿しているケースなどがあげられる。さらに、課題として、Twitter 上には、日々新しい隠語が生みだされているため、今後、新しい隠語のみで薬物売買が行われた場合、対応できない可能性が考えられる。抽出可能な投稿例を図 2、抽出できない投稿例を図 3 に示す。図 2 の投稿では、隠語や薬物の商品名が使用されているため、本システムにより抽出可能であった。しかし、図 3 の投稿に関しては、検出できなかった。図 3 の投稿では、投稿内容に新しい隠語が用いられており、さらに、この例は、画像に詳細情報を記載している。本論文では言語情報のみを解析しているため、現状では、画像内に含まれる隠語や商品名は検出に活用できない。3.2 節で実施した調査から、マイクロブログの特性上、フォロー/フォロワがハッシュタグ検索を用いてユーザを発見するため、多くの投稿については、



注) 網掛け部分は読者に有害な情報が含まれているため、掲載前にマスク処理を施した。

有害度判定に活用可能なキーワード

図 2 検出可能な投稿例

Fig. 2 Detectable post example.

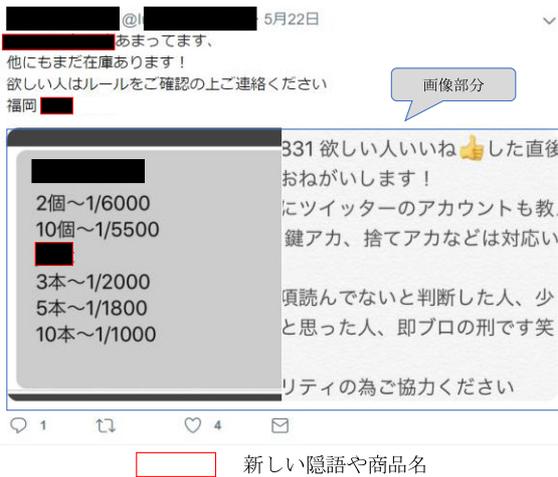


図 3 検出できない投稿例
 Fig. 3 Undetectable post example.

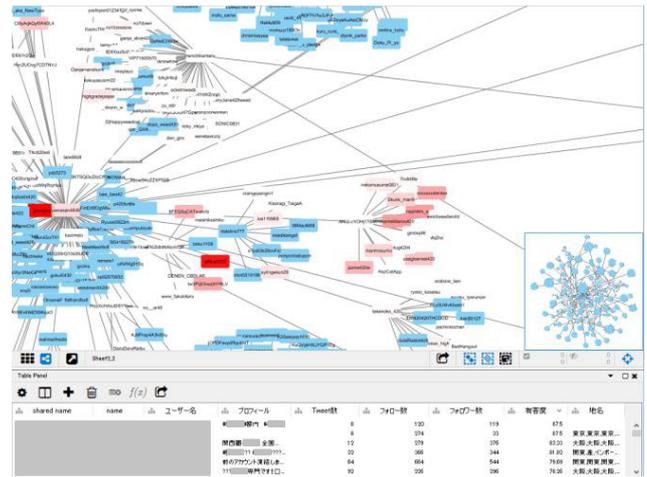


図 4 可視化結果
 Fig. 4 Result of Visualizing user network.

利用頻度の高い隠語と低い隠語が複合的に利用される。そのため、この課題については、本研究システムにより蓄積した有害な投稿内容から、新しい隠語を抽出し、有害単語の候補リストを自動的に抽出する仕組みを導入することで解決を目指していく。また、画像に関する対応は、有害画像の学習や OCR などによる抽出の実現可能性を検討していく。

本実験では、投稿ごとの有害判定精度を確認したが、実運用では、複数の投稿群からユーザの有害度を算出するため、本実験精度で十分実運用可能であると考ええる。

4.3 薬物に関するユーザネットワークの可視化
 4.3.1 実験内容

本実験では、有害度算出機能で判定したユーザごとの有害度を用いて、ユーザネットワークを可視化する。本研究では、可視化ツールに Cytoscape を採用した。これは、いかなる環境でも負担なくシステムを導入できると考えたためである。また、Cytoscape には、様々なアトリビュートを付随できるため、ユーザの情報を包括的に確認することが可能である。そこで、本システムでは、ユーザ判定 DB にユーザの有害度のみだけでなく、ユーザの投稿数、フォロワー数、MeCab を利用して取得した地名情報をあわせて蓄積し、Cytoscape を用いて可視化した。可視化結果のイメージを図 4 に示す。本システムでは、特定のユーザを確認するとユーザの ID や名前、プロフィール、投稿数、フォロー数、フォロワー数、有害度、地名の情報を確認できる。本システムでは、ユーザごとの有害度を簡易に見分けられるために色で判別する仕様とした。青色のユーザは無害なユーザや鍵アカウントなどの理由で判定できていないユーザ、赤色のユーザは有害なユーザである。有害度が高いほど濃い色を付与している。

本実験では、薬物売買を示唆する投稿を複数回行っているユーザを入力情報とし、本システムを用いてユーザ情報の

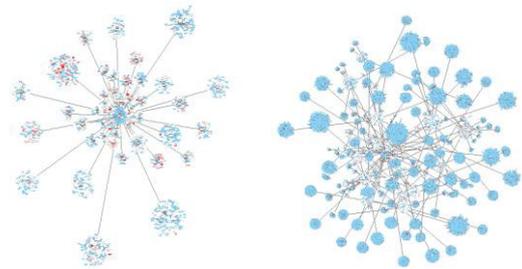


図 5 第 2 階層 (左), 第 3 階層 (右) の可視化結果
 Fig. 5 Result of Visualizing the second hierarchy (left) and the third hierarchy (right).



図 6 薬物売買コミュニティの可視化結果
 Fig. 6 Visualization result of drug dealing community.

可視化を行った。起点ユーザのフォロワ (第 1 階層: ユーザ数 89 件) を基点とし、そこから第 1 階層ユーザのフォロワネットワーク (第 2 階層: ユーザ数 2,901 件) と第 2 階層ユーザのフォロワネットワーク (第 3 階層: ユーザ数 14,957 件) を収集した。本データを対象にして分析する。

4.3.2 実験結果

第 2 階層と第 3 階層の可視化結果を図 5 に示す。この結果から、マイクロブログを対象に起点ユーザを入力情報として、ユーザのフォロワ間のつながりを可視化することが可能であることが分かった。さらにネットワークを詳細に確認することで、有害ユーザとそのフォロワのつながりや有害ユーザが集中している箇所を包括的に可視化できることが分かった。本システムにより可視化した薬物売買コミュニティの抜粋を図 6 に示す。図 6 より、ネットワー

あるといえる。実際に、県警のサイバーパトロール担当者から得た本システムに関するレビューは、以下のとおりである。

- サイバー空間では、違法・有害情報が氾濫しているほか、SNSなどのコミュニティサイトに起因する被害が増加している。そのような状況の中、私は、本システムを活用することにより、いち早く違法・有害情報を把握し、適切な対応が実現できると確信している。
- 危険と判断されるユーザのつながりを可視化できるこのシステムは、犯罪捜査において非常に有用な機能である。本システムを応用することで、薬物売買だけでなく、児童買春や自殺補助などの幅広い犯罪に活用可能であると考えている。

今後は、さらに高精度にユーザを抽出するため、リアルタイムに投稿を収集し、解析する仕組みを導入することを検討している。また、現段階のシステムでは、投稿記事だけを対象にしており、画像や絵文字は判定に活用していない。マイクロブログ上に画像が添付された薬物売買の投稿も多く、薬物を示唆する特定の絵文字なども特徴として取得可能なため、今後はこれらの情報も含めて解析することで、より高精度なシステムを実現できると考えている。

謝辞 本研究は、平成 29 年 12 月 26 日より、福井県警察本部からサイバー防犯ボランティアの委嘱を受け、県警と連携をとりながら研究開発を行ったものである。ヒアリングや教師データの作成にご協力いただいた県警の皆様、サイバー防犯ボランティアの皆様にご感謝の意を表す。また、本研究は、仁愛大学共同研究費と JSPS 科研費 19K14160 の助成を受けたものである。

参考文献

- [1] 警察庁：平成 29 年における SNS 等に起因する被害児童の現状と対策について (2018).
- [2] 株式会社ナナムエ：ひま部、入手先 (<https://himabu.com/>) (参照 2019-10-08).
- [3] 厚生労働省：第五次薬物乱用防止五か年戦略 (2018).
- [4] 警察庁：平成 29 年における SNS 等に起因する被害児童の現状と対策について (2018).
- [5] 中村健二, 山本雄平, 田中成典, 北野光一, 寺口敏生, 安彦智史, 辻 光宏：非行逸脱行為監視のためのネットパトロール支援システムの開発, フェジシステムシンポジウム講演論文集, Vol.27, pp.633-638, 日本知能情報フェジ学会 (2011).
- [6] 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, pp.571-584, 情報処理学会 (2013).
- [7] 大内祐起, 田中成典, 安彦智史, 若林克磨：見守りアプリケーションの開発, 第 75 回全国大会講演論文集, Vol.75, No.3, pp.111-112, 情報処理学会 (2013).
- [8] 山田 大, 安彦智史, 長谷川大, Michal Ptaszynski, 中村健二, 佐久田博司：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, 第 22 回年次大会発表論文集, pp.49-50, 言語処理学会 (2016).
- [9] 安彦智史, 長谷川大, Michal Ptaszynski, 中村健二, 佐久田博司：ID 交換掲示板における書き込みの隠語表記揺れを考慮した有害性評価, 情報システム学会誌, Vol.13, No.2, pp.41-58, 情報システム学会 (2018).
- [10] 安彦智史, 丸岡正朋, 長谷川大, 池辺正典：子供を取り巻くスマホアプリ～レビュー情報からみる危険性についての一考察, チャイルド・サイエンス (日本子ども学会論文誌), Vol.15, pp.44-48, 日本子ども学会 (2018).
- [11] Lee, W., Lee, S.S., Chung, S. and An, D.: Harmful Contents Classification Using the Harmful Word Filtering and SVM, *Proc. 7th International Conference on Computational Science*, pp.18-25, Springer-Verlag (2007).
- [12] 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-Verlag (2007).
- [13] Chandrinos, K.V., Androustopoulos, I., Paliouras, G. and Spyropoulos, C.D.: Automatic Web Rating: Filtering Obscene Content on the Web, *Proc. 4th European Conference on Research and Advanced Technology for Digital Libraries*, pp.403-406, Springer-Verlag (2000).
- [14] 松葉達明, 里見尚宏, 梶井文人, 河合敦夫, 井須尚紀：学校非公式サイトにおける有害情報検出, 言語理解とコミュニケーション研究会技術研究報告, Vol.109, No.142, pp.93-98, 電子情報通信学会 (2009).
- [15] 菊池琢弥, 内海 彰：語の共起情報に基づく有害サイトフィルタリング手法, 第 9 回情報科学技術フォーラム講演論文集, Vol.9, No.2, pp.1-6, 情報処理学会・電子情報通信学会 (2010).
- [16] 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, DEIM Forum 2010, 日本データベース学会 (2010).
- [17] 池田和史, 柳原 正, 服部 元, 松本一則, 小野智弘, 滝嶋康弘：HTML 要素に基づく有害サイト検出手法, 情報処理学会論文誌, Vol.52, No.8, pp.2474-2483, 情報処理学会 (2011).
- [18] Breiman, L.: Random Forests, *Machine Learning*, Vol.45, No.1, pp.5-32 (2001).
- [19] Cortes, C. and Vapnik, V.: Support-Vector Networks, *Machine learning*, Vol.20, No.3, pp.273-297, Springer-Verlag (1995).
- [20] Ribeiro, M.H., Calais, P.H., Santos, Y.A., Almeida, V.A.F. and Meira, W. Jr.: Characterizing and Detecting Hateful Users on Twitter, arXiv preprint arXiv:1803.08977 (2018).
- [21] Al-garadi, M.A., Varathan, K.D. and Ravana, S.D.: Cybercrime Detection in Online Communications: The Experimental Case of Cyberbullying Detection in the Twitter Network, *Proc. Computers in Human Behavior*, Vol.63, pp.433-443 (2016).
- [22] Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp.230-237 (2004).
- [23] Ptaszynski, M., Eronen, J.K.K. and Masui, F.: Learning Deep on Cyberbullying is Always Better Than Brute Force, *Proc. Linguistic and Cognitive Approaches to Dialogue Agents (LaCATODA 2017), CEUR Workshop Proceedings*, Vol.1926, pp.3-10 (2017).
- [24] Twitter: Twitter Developer, available from (<https://developer.twitter.com/>) (accessed 2019-10-08).



安彦 智史 (正会員)

1985年生。2008年関西大学総合情報学部総合情報学科卒業。2010年関西大学大学院総合情報学研究科博士課程前期課程修了。2013年関西大学大学院総合情報学研究科博士課程後期課程修了。博士(情報学)。同年青山学院大学附置情報メディアセンター助手。2016年仁愛大学人間学部コミュニケーション学科講師に着任し、現在に至る。画像処理, Webマイニング, 子ども学に関連する研究に従事。情報システム学会, 土木学会等各会員。



加藤 諒 (正会員)

1989年生。2012年関西大学総合情報学部総合情報学科卒業。2014年関西大学大学院総合情報学研究科博士課程前期課程修了。2017年関西大学大学院総合情報学研究科博士課程後期課程修了。博士(情報学)。2012年(株)関西総合情報研究所入社。システム設計等の研究開発に従事。2017年京都橘大学現代ビジネス学部助教, 現在に至る。知識情報処理, Webマイニング等の研究に従事。電子情報通信学会会員。



北川 悦司 (正会員)

1977年生。2000年関西大学総合情報学部総合情報学科卒業。2002年関西大学大学院総合情報学研究科博士課程前期課程修了。2005年関西大学大学院総合情報学研究科博士課程後期課程修了。同年博士(情報学)を取得, 阪南大学経営情報学部専任講師。2007年准教授。2013年教授, 現在に至る。2014年4月から1年間カナダブリティッシュコロンビア大学(UBC)客員准教授。写真測量や画像処理, 3次元点群処理に関する研究に従事。2000年(株)関西総合情報研究所を起業, 2001年から2013年まで取締役(代表取締役の期間もあり)。3次元処理, CAD/CG, GIS/GPS, 画像処理, そしてWebソリューションビジネスに関連する研究業務に従事。情報知識学会, 土木学会, 日本知能情報ファジィ学会各会員。