

# SNS 上の違法薬物売買に関する投稿記事の検出手法の開発

佐藤 大宙<sup>†</sup> 村上 洋一<sup>‡</sup>東京情報大学総合情報学研究科<sup>†</sup> 東京情報大学総合情報学部<sup>‡</sup>

## 1. はじめに

薬物事犯の検挙人員は、近年、減少傾向が続いている[1]。例えば、覚醒剤事犯の検挙人員は第三次覚醒剤乱用期のピークであった平成9年の19722人から長期的に減少し、令和4年は6124人であった。また令和4年の大麻事犯の検挙人員は、過去最多であった前年の6900人を下回り、5342人であった。しかしながら、依然として薬物事犯が発生しており、引き続き、厳正な取締りに加え、若年層による乱用防止を主な目的として、インターネット上での違法情報・有害情報の排除や広報啓発活動が推進されている[2]。特に、近年、スマートフォンの普及等により、手軽にインターネットを利用できる環境となったことで、匿名性の高い SNS を利用した薬物密売が行われるなど密売・購入手法について潜在化・巧妙化が一層進行していると言われている[2]。

そこで本研究では、SNS 上の違法薬物売買に関する投稿記事の検出を目指し、投稿記事で使用される隠語の有無や対数オッズ比、また単語の一般的な使われ方であるかを考慮するスコア（一般性スコア）を用いた分類手法の開発する。また、既知のすべての隠語を検索語として用いて投稿記事を収集し、人手によるラベル付けを行い、信頼性の高いデータセットを作成する。

## 2. 関連研究

安彦ら[3]は、機械学習を用いて薬物売買を行うユーザを自動抽出する手法を提案した。人手により分類して作成した違法薬物に関する投稿記事 2094 件、それ以外の投稿記事 5870 件を用いて、それらの投稿記事に含まれる名詞単語の出現頻度から特徴ベクトルを作成し、学習を行い、有害判定モデルを構築した。

10 分割交差検証により、投稿記事ごとの有害判定の精度として正解率を算出した結果、Random Forests (RF) の場合、0.933 を達成した

と報告している。松波ら[4]は、絵文字にも着目し、違法薬物に関する投稿記事を分類する手法を提案した。隠語 4 語を用いて収集した投稿記事を、人手により分類した違法薬物に関する投稿記事 2941 件、それ以外の投稿記事 4598 件を用いて、名詞、動詞、形容詞、副詞の単語や記号の TF-IDF 値だけでなく、絵文字に関する 4 つの特徴: ①絵文字の出現頻度、②薬物隠語絵文字の出現頻度+果物絵文字の有無、③絵文字ごとの出現頻度、④薬物隠語絵文字と果物絵文字ごとの出現頻度、を追加して学習を行った。10 分割交差検証の結果、絵文字の特徴③と④を用いた LightGBM の場合、適合率 0.984、再現率 0.954、F 値 0.969 を達成したと報告している。

一方、青木ら[5]は、SNS において辞書に掲載されている用法とは全く異なる使われ方がされている単語に着目し、単語の一般的ではない用法を検出する手法を提案した。これは着目単語とその周辺単語の単語ベクトルを利用し、周辺単語が、均衡コーパスにおける一般的な用法の場合の周辺単語とどの程度異なっているかを評価して、着目単語の一般性スコアが算出される。このスコアが高い場合に一般的である用法、低い場合に一般的ではない用法と判断される。

## 3. 違法薬物売買に関する投稿記事の検出手法

### 3.1. 違法薬物に関する隠語を含む投稿記事の収集と学習データセットの作成

全国の警察、警察庁、警視庁のホームページで紹介されている違法薬物に関する隠語のうち、名詞であると判断した 177 種類の隠語を取得し、投稿記事の収集の検索語として用いる。

収集された投稿記事の文について、以下の整形処理を行う。

- ID, 文の重複を削除
- 半角カタカナを全角に変換
- 大麻の隠語「420」を除く数字について桁数を保持しないで「0」に変換
- MDMA の隠語「x」を除くすべての記号を削除
- URL, エスケープ処理処理用の文字列（「&lt;」 「&gt;」 など）を削除
- 絵文字の性別、絵文字の肌の色、絵文字の髪

Development of a method for detecting posted articles about illegal drug sales on SNS

<sup>†</sup> Daichi Sato, Graduate School of Informatics, Tokyo University of Information Sciences

<sup>‡</sup> Yoichi Murakami, Faculty of Informatics, Tokyo University of Information Sciences

型を削除

- イングランド、スコットランド、ウェールズの旗をイギリスの旗に変換

次に、投稿記事が違法薬物に関するものか（悪性）、それ以外か（良性）の人手によるラベル付を行う。その後、著者によるラベル付けを行い、両者のラベル付けが一致した投稿記事のみをデータセットとして採用する。

### 3.2. 対数オッズ比の計算

ある単語について、悪性データセットと良性データセットそれぞれに含まれる確率と含まれない確率の比をとり、悪性と良性の比に対数をとる。  $P_{pos}$  を悪性データセットでの出現確率、  $P_{neg}$  を良性データセットでの出現確率とするとき、次の数式で表される:

$$\log(odds) = \log\left(\frac{P_{pos}/1-P_{pos}}{P_{neg}/1-P_{neg}}\right)$$

### 3.3. 一般性スコアの計算

一般性スコア[5]は、着目単語の入力側ベクトルとその周辺単語の出力側ベクトルの内積にシグモイド関数を適用し、その加重平均をとることで計算する。単語ベクトルはライブラリ Gensim の Word2Vec を利用し、Wikiedia のデータを文書集合として使用することで生成する。WikiExtractor を利用して Wikipedia の XML データをテキストデータに変換し、ID の重複削除、文の重複削除以外のものに加え、<doc>タグの削除を行う。

### 3.5. 評価実験

次の3つの特徴量を用いて分類器を作成し、10分割交差検証により性能評価を行う。

- ① 隠語の有無のみを用いた分類器
- ② 隠語の対数オッズ比のみを用いた分類器
- ③ 隠語の一般性スコアのみを用いた分類器

機械学習法は、安彦ら[3]と同様に、Support Vector Machine (SVM), Random Forest (RF), Navive Bayes Classifier (NBC), Light Gradient Boosting Machine(LightGBM)を用いる。

### 4. 性能評価と考察

177 種類の隠語を検索語として収集した投稿記事に対して人手によるラベル付を行い、悪性の投稿記事 3472 件、良性の投稿記事 17499 件のデータセットを作成した。隠語の含む間バイトに関する投稿記事については、データセットには含めなかった。

各分類器の評価実験の結果を表 1-3 に示す。い

ずれの分類器においても、NBC の除く、SVM, RF, LightGBM を用いた分類器は、先行研究よりも高い性能を示すことが確認できた。

しかしながら、それぞれの分類器では検出できなかった投稿記事（偽陰性）が存在していた。例えば、押しや対面取引の隠語である絵文字「❤️」のみが含まれている投稿記事は、正しく検出することができなかった。

表 1. 隠語の有無のみを用いた分類器

	SVM	RF	NBC	LightGBM
適合率	0.998	0.999	0.215	0.999
再現率	0.985	0.986	0.998	0.986
F 値	0.991	0.992	0.354	0.992
ROC AUC	0.998	0.999	0.639	0.999
PR AUC	0.997	0.998	0.607	0.998

表 2. 隠語の対数オッズ比のみを用いた分類器

	SVM	RF	NBC	LightGBM
適合率	0.998	0.999	0.215	0.999
再現率	0.985	0.986	0.998	0.986
F 値	0.991	0.992	0.354	0.992
ROC AUC	0.999	0.999	0.639	0.999
PR AUC	0.997	0.998	0.607	0.998

表 3. 隠語の一般性スコアのみを用いた分類器

	SVM	RF	NBC	LightGBM
適合率	0.999	0.993	0.226	0.995
再現率	0.949	0.973	0.996	0.972
F 値	0.973	0.983	0.368	0.983
ROC AUC	0.987	0.996	0.662	0.996
PR AUC	0.982	0.992	0.612	0.992

### 5. おわりに

本研究では、違法薬物に関する投稿記事を、3つの特徴量を用いて分類する手法について、人手によってラベル付されたデータセットを用いて評価実験を行った。その結果、先行研究よりも高い性能を示すことが確認できた。今後は、各種特徴量を組み合わせ、また隠語以外の特徴量も検討を行い、未知の隠語を含む投稿記事の分類に対応できる分類器を開発する。

### 参考文献

- [1]警察庁. "組織犯罪対策に関する統計".
- [2]厚生労働省. "薬物乱用対策".
- [3]安彦智史他. 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. 2020, vol.61, no.3, pp.535-543.
- [4]松波琴未他. Twitter からの違法薬物売買ユーザの抽出. 情報処理学会第 83 回全国大会講演論文集. 2021, vol.1, no.1, pp.533-534.
- [5]青木竜哉他. ソーシャルメディアにおける単語の一般的ではない用法の検出. 自然言語処理. 2019, vol.26, no.2, pp.381-406.