

位置および投稿者希少性に着目した 犯罪関連投稿の抽出手法

高橋 柊^{1,a)} 菊地 悠¹ 深澤 佑介¹

概要：近年，Twitter 等のマイクロブログに代表される SNS サービスのユーザ数増加およびスマートフォンの普及により，実世界における情報がリアルタイムに web 上にアップロードされるようになった．そのため，マイクロブログ上の情報をセンシングすることで，実世界における事象を検知する研究が活発となっている．本研究では，マイクロブログ上に投稿される犯罪関連投稿に着目する．マイクロブログより，リアルタイムな犯罪関連投稿を抽出することが出来れば，犯罪事象に対し短時間で適切な防犯対策が可能となる．提案手法により，ユーザの投稿内容，投稿関連位置情報および，関連ステータス情報を用いることで，マイクロブログより希少性のある犯罪関連投稿を抽出することが可能となる．Twitter の日本語投稿データに対し提案手法を適応したところ，辞書集合とのマッチング抽出に対し，希少性が高い犯罪関連情報の抽出が可能となる事が示唆された．

Extraction of Criminal Blogs based on Scarcity of Locations and Users

TAKAHASHI SHU^{1,a)} KIKUCHI HARUKA¹ FUKAZAWA YUSUKE¹

1. はじめに

近年，ソーシャルメディア上のデータを用いた社会センシング技術の研究が活発化している．リアルタイムに市民が情報を投稿するソーシャルメディアは，既存のセンシングデバイスでは抽出困難なリアルタイム情報を多く含むため，社会的な事象が大きく反映される．これらの研究は Human Probe(HPB) と呼ばれ，社会およびビジネス分野への応用が進んでいる．

ソーシャルメディアを用いた Human Probe 研究において，イベント情報を検出する手法が多く提案されている．Chen ら [1] は画像投稿コミュニティサイトである Flickr^{*1} の投稿情報を用いることで，イベントを検出する手法について検討している．しかし，投稿数の出現回数に依存するため，ごく少数の投稿からイベントを抽出することは困難

である．山田ら [2] は，Twitter^{*2} からイベント情報を自動的に抽出する技術について検討している．Twitter を情報源として活用することで，多様かつ大量なイベント情報の抽出が可能となることを示している．

本研究では，Twitter に投稿される犯罪関連投稿の抽出に焦点をあて，位置および投稿者希少性を用いた投稿抽出手法について検討する．Twitter より投稿者が経験あるいは目撃した犯罪関連投稿を抽出することができれば，警察によるパトロールなど既存のセンシングでは顕在化困難であった犯罪事象についての知見を得ることができる．

位置および犯罪事象との関連性を分析する Geographic profiling [3] は，犯罪の防止および捜査に利用されている．Geographic profiling により，犯罪の空間的な動作を把握し，重点的に治安維持活動を行う地域を特定することができる．また，ソフトウェアを用いることで，多量のデータから効率的に Geographic profiling を行うアルゴリズムおよびシステムが提案されている [4][5]．

¹ 株式会社 NTT ドコモ
NTT DOCOMO, INC., Yokosuka, Kanagawa 239-8536, Japan

^{a)} syuu.takahashi.us@nttdocomo.com

^{*1} Flickr, <https://www.flickr.com/>

^{*2} Twitter, <https://twitter.com>

Geographic profiling では、一般に捜査活動より得られたデータや過去の犯罪統計情報を利用することで位置および犯罪事象との関連性を分析する。しかし、ソーシャルメディア上のデータより犯罪関連情報を抽出することができれば、より多様なデータから高度な Geographic profiling が可能となる。White ら [6] は Twitter 投稿情報及びソーシャルグラフを用いた Geographic profiling の有効性について検討している。あらかじめ設定された犯罪行為者のソーシャルグラフを用いた分析や、キーワードの出現数変化を捉えることで、Twitter が Geographic profiling に有効であることを示している。また、Wang ら [7] は LDA および線形モデルを用いることで、Twitter の投稿情報からひき逃げ事件の予測が可能であることを示している。しかし、既存研究では、あらかじめ与えられた特定の犯罪情報を Twitter から抽出するため、多様な犯罪関連投稿を抽出することは困難である。

本研究では、多様な犯罪関連投稿を Twitter から抽出する手法について検討する。日本において市民が犯罪にあふ経験はごく希少な確率で発生すると考えられる。提案手法ではごく希少な確率で発生するという犯罪の特性を利用することで、犯罪関連投稿を抽出する。Twitter はリアルタイム性の高いソーシャルメディアであり、投稿者の投稿頻度が他のソーシャルメディアと比べ高い。そのため、犯罪事象が起こることで、投稿者の投稿内容あるいは位置における語の出現分布に違いが起こることが期待できる。

Twitter に投稿されるツイートには、多くの場合位置情報が付与されていない。そこで、投稿内容に関連した位置情報を新たに付与する必要がある。投稿者が犯罪を経験あるいは目撃したことについて投稿した場合、投稿テキスト中に投稿関連位置に関する語が出現することが期待できる。本研究では、落合ら [8] による投稿テキストと投稿関連位置を紐付ける手法を利用する。

2. 提案手法

提案手法では、Twitter の投稿情報より投稿者が犯罪を経験あるいは目撃した投稿を抽出することで、警察官のパトロールなど既存のセンシングでは顕在化が困難であった犯罪関連投稿を抽出する。図 1 に提案手法における分析ステップを示す。

本提案手法は以下 4 つの分析ステップにより構成される。

i) 犯罪語マッチ処理

Twitter より取得した投稿および、犯罪関連語を含む犯罪語辞書を用い、投稿テキスト中に犯罪語辞書の単語を含む投稿を抽出する。

ii) 投稿者希少性を考慮した特徴量計算

犯罪語マッチより得られた投稿の投稿者が、過去に同様の犯罪事象について投稿しているか否かを計算する。

iii) 位置希少性を考慮した特徴量計算

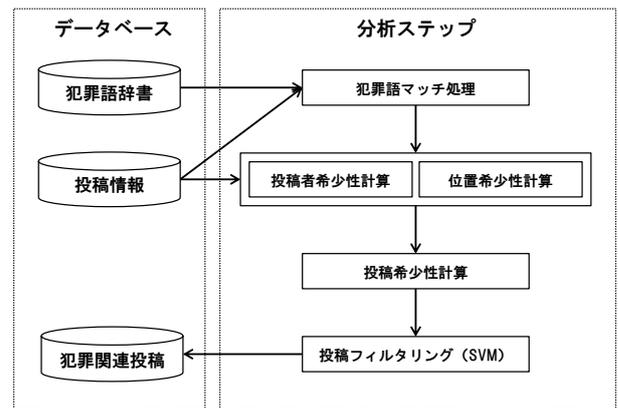


図 1 分析ステップ

犯罪語マッチより得られた投稿における関連位置において、過去に類似した投稿がされているか否かを計算する。

iv) 投稿フィルタリング

投稿者および位置希少性により計算された特徴量および、投稿テキストより抽出可能な特徴量を利用し、犯罪関連投稿を抽出する分類器を構築する。

2.1 犯罪語マッチ処理

投稿内容が犯罪関連語を含んでいる投稿を抽出したい。投稿テキスト中に犯罪関連語を含むか否かを判定するために、犯罪語辞書を作成した。表 1 に作成した犯罪語辞書を示す。日本において強盗や殺人などの重犯罪は、注目度が高く、警察庁などが統計情報を公表している。本研究では、ソーシャルメディア特有の犯罪関連投稿を抽出したい。そのため、比較的刑事罰が弱く、また通報されることが少ないと考えられる犯罪を対象し、犯罪語辞書を構築した。

表 1 犯罪語辞書

犯罪語	ルビ
痴漢	チカン
ナンパ	ナンパ
覗き	ノゾキ
置き引き	オキビキ
キャッチ	キャッチ
盗撮	トウサツ
泥棒	ドロボウ
万引き	マンビキ
ストーカー	ストーカー
ひったくり	ヒッタクリ
酔っ払い	ヨッパライ
変質者	ヘンシツシャ
警官	ケイカン
警察	ケイサツ
騒ぎ	サワギ
スカウト	スカウト
セールス	セールス
パトカー	パトカー

投稿情報および、犯罪語辞書を用い、犯罪関連語を含む投稿情報のみを抽出する。Twitterの投稿テキストでは、一般に漢字で表記されるような語についても、ひらがなやカナカナを用いて投稿されるなど、表記ゆれが多分に見受けられる。

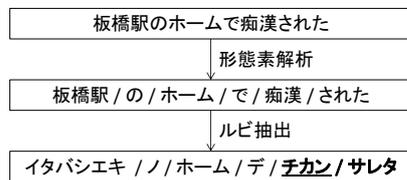


図2 キーワードマッチングの実施例

図2にキーワードマッチングの実施例を示す。例えば、「痴漢、ちかん、チカン」はすべて同様の犯罪事象について投稿しているが、「痴漢」のみをキーワードマッチングの条件とすると、「ちかん、チカン」については、除外されてしまうという問題がある。そこで、投稿テキストをJTAG[9]を用い形態素解析し、各形態素をルビに変換したのち、キーワードマッチングを行う。

2.2 投稿者希少性を考慮した特徴量計算

日本において、同一人物が数ヶ月間に複数回犯罪を経験あるいは目撃することは少ない。そのため、あるユーザが複数回犯罪に関する投稿をしている場合、それはユーザの経験あるいは目撃した投稿ではなく、ニュースなどの伝聞情報やノイズである可能性が高い。例えば、「泥棒」という単語を高頻度で投稿している投稿者は、泥棒を経験あるいは目撃しているのではなく、ゲームや映画など非現実世界における事象について投稿している可能性が高い。また、「警察」という単語を高頻度で投稿している投稿者は、警察を目撃したのではなく、警察関連組織における、公式アカウントである可能性が高い。一方で、普段「泥棒」という単語を投稿していない投稿者が「泥棒」という単語を投稿した場合、投稿者が泥棒を経験あるいは目撃した可能性が高いといえる。

そこで、ある犯罪事象に関連した投稿をしていた投稿者が一定期間中に同様の事象について投稿しているか否かを投稿者希少性スコアとして計算する。投稿者が過去同様の犯罪事象に関して投稿していないほど、投稿内容は投稿者の経験あるいは目撃した内容である可能性が高いといえる。

投稿者希少性は、 $tf \cdot idf$ 法を用いて計算する。犯罪語マッチより得られた投稿を対象投稿とする。対象投稿における投稿者の過去投稿を抽出する。抽出された全投稿数を N 、対象投稿 d における犯罪語 T の出現頻度を $tf(d, T)$ 、犯罪語 T を含む投稿数を $df(T)$ とした時、投稿者希少性 $user_score(T, d)$ は式1より計算される。

$$user_score(T, d) = tf(d, T) \cdot \log \frac{N}{df(T)} \quad (1)$$

2.3 位置希少性を考慮した特徴量計算

任意の場所においてある犯罪事象に関する投稿が過去多く検出される場合、その情報は社会的に知られている可能性が高く、また犯罪関連投稿と考えずらい。例えば、「痴漢」という単語がある任意の地域において過去多く投稿されている場合、「痴漢」が恒常的に発生しているか、「痴漢」に関連する犯罪事象以外の事象が恒常的に発生している可能性が高いといえる。このように、ある地域における恒常的な事象は、地域における投稿中に含まれる語の分布によって示される。

位置希少性は、犯罪語マッチより得られた対象投稿における投稿位置および投稿日より、投稿位置と紐付いた、投稿日中の投稿と過去の投稿を比較することで計算する。投稿日の投稿群を犯罪日投稿群、過去の投稿を過去投稿群とする。犯罪日投稿群における単語の出現率が、過去投稿群における単語の出現率と異なるほど、犯罪日における位置希少性が高いといえる。

犯罪日投稿群および、過去投稿群の単語の出現率を確率分布とし、Jensen-Shannon divergenceを用い確率分布の類似度を計算する。犯罪日投稿群における単語の出現分布を P 、過去投稿群における単語の出現分布を Q としたとき、犯罪日希少性 $poi_score(P, Q)$ は式2より計算される。

$$poi_score(P, Q) = D_{JS}(P||Q) \quad (2)$$

$$\text{ただし, } D_{JS}(K||L) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M)$$

$$M = \frac{1}{2}(P + Q)$$

$$D_{KL}(P||M) = \sum_i \log_2 P(i) \frac{P(i)}{M(i)}$$

2.4 投稿フィルタリング

文章分類問題において、テキストをBag-of-Wordsで表現し、分類器を構築する手法が広く使われている。Dilrukshira[10]はTwitterの投稿テキストをラベリングし、Support Vector Machine (SVM)を利用することで、投稿テキストを分類する手法について検討している。SVMを利用することで多量の特徴量を評価した分類器の構築が可能となることを示している。

本手法では、SVMを用い犯罪関連投稿を抽出する分類器を構築する。多量の投稿テキストをBag-of-Wordsで表現すると、次元数は数万に及ぶ。次元数が膨大になることで、計算コストが増大してしまうという問題がある。そのため、Probabilistic Latent Semantic Analysis(pLSA)[11]を用いて、次元数を300次元に縮約する。

SVMを用いて分類器を構築する際、特徴量に位置および投稿者希少性を考慮した特徴量である poi_score および $user_score$ を追加することで、位置および投稿者希少性を考慮した犯罪関連投稿抽出を行う。

3. 評価

Twitter より提供されている, Public streams^{*3} を用い, Twitter の日本語投稿データを取得した. データ取得期間は 2015 年 1 月 1 日から 2015 年 9 月 31 日までの 9 ヶ月間であり, 日本語投稿および, Twitter 公式のクライアントを利用して投稿されたツイートを対象とした. 犯罪語辞書を用いたキーワードマッチングにて, 犯罪語を含む日本語ツイートのうち, 3600 件をランダムサンプリングし, 犯罪関連投稿である投稿を *True*, それ以外を *False* としラベリングした.

ラベリングした結果, 3600 件中 *True* が 713 件, *False* が 2887 件となった. ラベル付きデータを評価対象データとし, 投稿者希少性および位置希少性について計算することで, 提案手法の有用性について評価する.

3.1 投稿者希少性を考慮した特徴量

評価対象データに対し, *user_score* を計算した. 評価対象データにおいて, *user_score* は平均 0.242, 分散 0.100, 中央値 0.221 となった. 図 3 に *user_score* の出現度数を示す. 多くの投稿は *user_score* が低く, 少数の投稿のみ *user_score* が高いことがわかる.

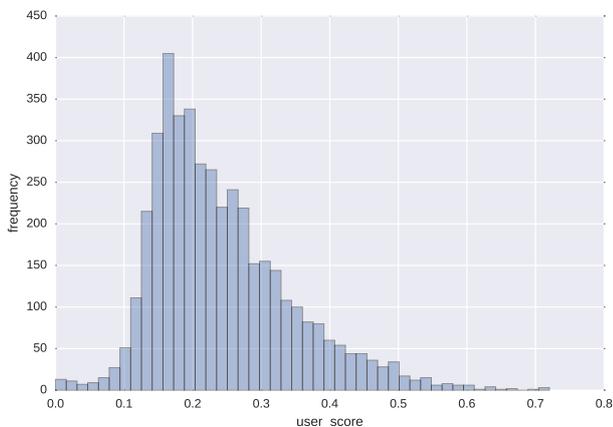


図 3 *user_score* の出現度数

図 4 に, ラベリング結果と *user_score* の分布比較を示す. *label = False* の投稿では, *user_score* が平均 0.229, 中央値が 0.208 に対し, *label = True* の投稿では, 平均 0.294, 中央値 0.283 となった. 犯罪関連投稿では, 非犯罪関連投稿と比較して *user_score* が高くなる傾向がある. したがって, *user_score* の値が犯罪関連投稿を抽出する際に有益な値であるといえる.

表 2 に *user_score* が高い投稿および低い投稿 5 件における投稿テキストの一部を示す. *user_score* が高い投稿は, 犯罪関連投稿が多く, *user_score* が低い投稿は, 広告や定常的な投稿であることがわかる. 特に, 警察署の広報アカ

*3 Public streams, <https://dev.twitter.com/streaming/public>

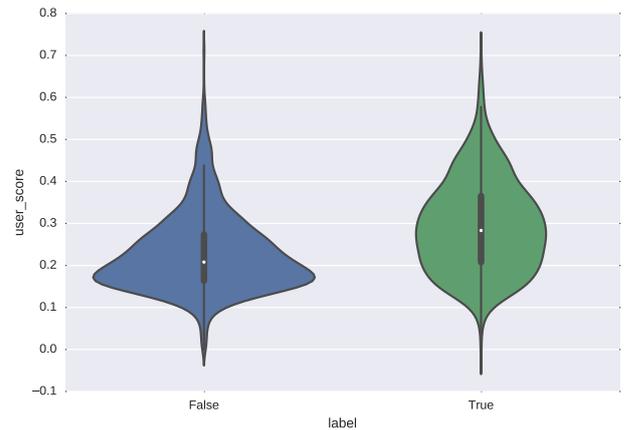


図 4 ラベルにおける *user_score* の分布

ウント等による投稿は, *user_score* が低くなる傾向が確認された.

表 2 *user_score* と投稿内容

投稿テキスト	犯罪語	score	label
厚木で酔っ払い。	酔っ払い	0.720	True
博多駅パトカー	パトカー	0.707	True
ストーカー大川	ストーカー	0.707	False
御茶ノ水の酔っ払い	酔っ払い	0.696	False
なんぼだなんぼ	ナンパ	0.663	True
スカウトマン募集!勤務地...	スカウト	0.000	False
山陽小野田警察署 安全...	警察	0.000	False
器・盗撮器発見サービス」...	盗撮	0.000	False
ららぽーと横浜に盗撮魔が...	盗撮	0.000	True
下川町?集団ストーカー?...	ストーカー	0.001	False

3.2 位置希少性を考慮した特徴量

評価対象データに対し, *poi_score* を計算した. 評価対象データにおいて, *poi_score* は平均 0.492, 分散 0.090, 中央値 0.479 となった. 図 5 に *poi_score* の出現度数を示す. *user_score* の分布に偏りがあったのに対し, *poi_score* は正規分布に近い分布を示していることがわかる.

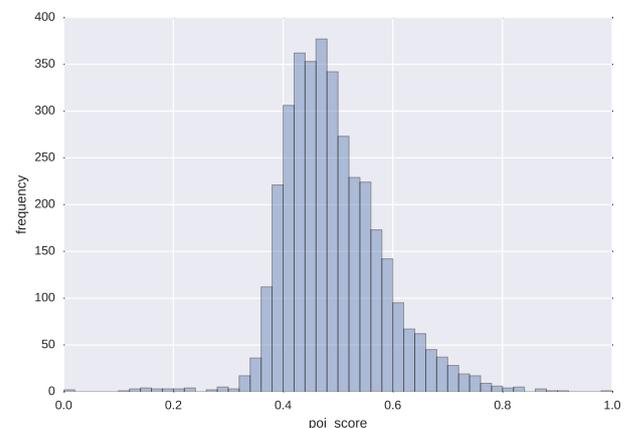


図 5 *poi_score* の出現度数

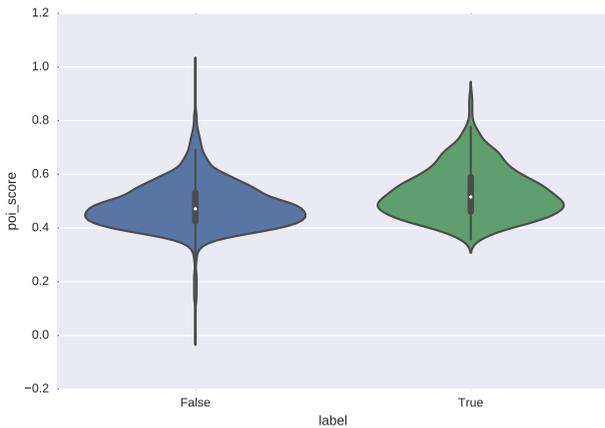


図 6 ラベルにおける *poi_score* の分布

図 6 に、ラベリング結果と *poi_score* の分布比較を示す。*label = False* の投稿では、*poi_score* が平均 0.482，中央値が 0.472 に対し，*label = True* の投稿では，平均 0.530，中央値 0.516 となった。犯罪関連投稿では，非犯罪関連投稿と比較して *poi_score* が高くなる傾向がある。したがって，*poi_score* の値が犯罪関連投稿を抽出する際に有益な値であるといえる。

表 3 に *poi_score* が高い投稿と低い投稿を 5 件示す。*poi_score* が高い投稿は，犯罪関連投稿が多く，*poi_score* が低い投稿は，広告や定常的な投稿であることがわかる。また，*poi_score* による投稿情報のソートを行うと，*user_score* によるソートと類似した結果が得られることが確認された。

表 3 *poi_score* と投稿内容

投稿テキスト	犯罪語	score	label
博多駅パトカー	パトカー	0.999	True
ストーカー大川	ストーカー	0.909	False
スキノキャッチやびゃあ	キャッチ	0.895	True
池袋=なんば	ナンバ	0.879	True
ストーカー乙変態乙乙	ストーカー	0.869	False
器・盗撮器発見サービス」...	盗撮	0.000	False
上尾警察で保護埼玉県鴻巣保...	警察	0.000	False
(不審者情報)[件名] 痴漢...	痴漢	0.114	False
盗撮...風紀乱す行為絶えず...	警察	0.129	False
(不審者情報)[件名] 痴漢事...	痴漢	0.137	False

3.3 投稿フィルタリング

既存手法における投稿テキストの Bag-of-words のみを利用した分類器の精度をベースラインとし，特微量として位置および投稿者希少性を追加した分類器の精度を比較することで，位置および投稿者希少性の有用性について評価する。

図 7 に評価対象より計算された *poi_score* および *user_score* の分布を示す。*poi_score* と *user_score* には正の相関があることがわかる。*poi_score* が高いほど，*user_score*

も高い値を示すことから，*poi_score* と *user_score* はともに，投稿の希少性を示しているといえる。希少性が犯罪関連投稿の抽出に有効であることを確かめるために，*poi_score* と *user_score* を特微量として追加した分類器の精度を本手法の評価値とする。

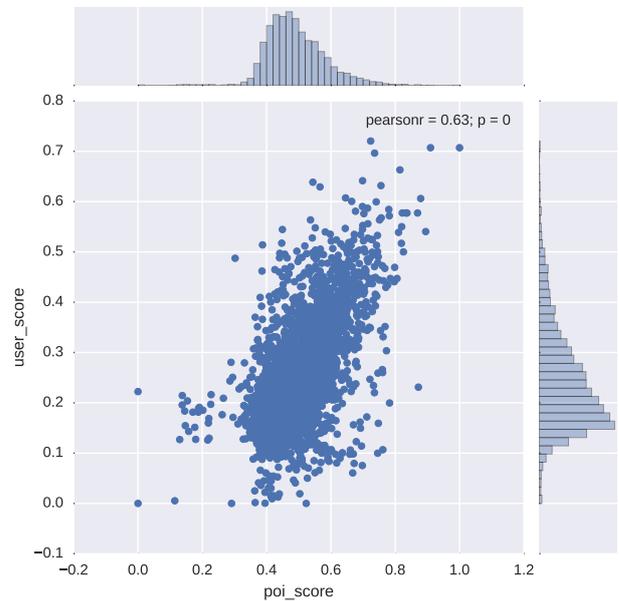


図 7 *poi_score* および *user_score* の分布

評価対象データのうち，ランダムにサンプリングした 50% の投稿 (1800 件) を訓練データとし，残りの投稿をテストデータとした。訓練データより構築された分類器の精度比較には，テストデータを分類した際の *precision*，*recall* および *F1score* を用いる。

表 4 に，分類器精度を示す。ベースラインでは，分類器の精度を示す *F1score* は 0.72 と比較的高いが，犯罪関連投稿の抽出性能を示す *label = True* では，0.47 とうまく分類ができていないことがわかる。一方，提案手法における位置および投稿者希少性を特微量として追加した場合，ベースラインと比べ，モデル全体の *F1score* が向上し，また *label = True* における *precision* および *recall* も向上していることがわかる。

表 4 学習器のスコア

	label	precision	recall	F1	support
baseline	False	0.89	0.70	0.79	1433
	True	0.37	0.67	0.47	367
	(avg)	0.79	0.70	0.72	1800
work	False	0.90	0.71	0.80	1433
	True	0.39	0.71	0.50	367
	(avg)	0.80	0.71	0.74	1800

図 8 にベースラインと提案手法における分類器精度の ROC カーブを示す。ROC カーブ下における面積 (AUC) が大きいほど，分類器がうまく犯罪関連投稿を抽出してい

る。ベースラインにおける AUC が 0.7367 なのに対し、提案手法では 0.7656 となった。提案手法を特徴量として追加した分類器の AUC をうわまっていることから、提案手法における位置および投稿者希少性が犯罪関連投稿の抽出に有用であることがわかった。

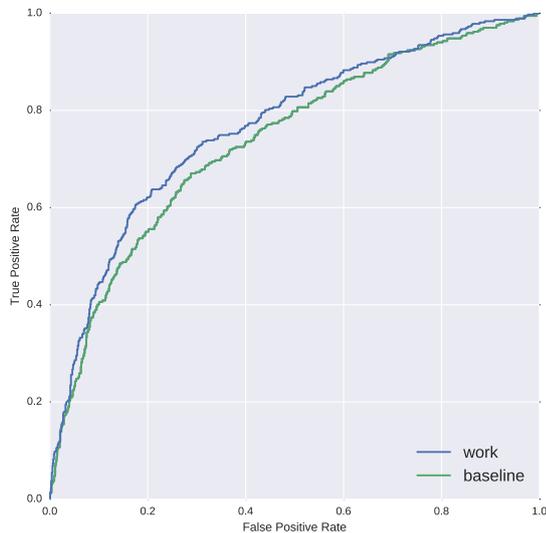


図 8 ROC カーブ

4. まとめ

本研究では、位置および投稿者希少性を用いることで、Twitter より犯罪関連投稿を抽出する手法について検討した。犯罪関連投稿と非犯罪関連投稿において、位置および投稿者希少性を比較した際、分布に差が出ることを示した。

評価対象データを bag-of-words で表現したのち、pLSA を用い次元縮約を行い、SVM を用いて犯罪投稿および非犯罪投稿を分類する分類器を構築した。SVM の特徴量に bag-of-words のみを用いた従来手法における分類精度と、提案手法を取り入れた分類精度を比較することで、提案手法の有用性について評価した。

今後の課題として、分類器の精度を向上させ、犯罪関連投稿抽出時の precision および recall を向上させることが挙げられる。犯罪関連投稿と非関連投稿とを比較した際、位置および投稿者希少性の分布がより異なれば、分類器の精度は向上するといえる。そのため、犯罪関連投稿とより関連のある希少性あるいは他の指標について検討する。また、キーワードマッチングに利用する犯罪関連語について、任意に与えた語集合に対し集合拡張を行うことで、より多様な犯罪に対応した犯罪語辞書の拡充が可能となる。そのため、集合拡張を用いることで犯罪語を半自動的にソーシャルメディア上のデータから抽出する手法についても検討する。

参考文献

- [1] Chen, L. and Roy, A.: Event Detection from Flickr Data Through Wavelet-based Spatial Analysis, *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09*, New York, NY, USA, ACM, pp. 523–532 (2009).
- [2] 山田, 悠 菊地, 桂一落合, 大祐鳥居, 浩 稲村, 賢 太田: マイクロブログを用いたイベント情報抽出技術, *情報処理学会論文誌*, Vol. 57, No. 1, pp. 123–132 (2016).
- [3] Harries, K.: *Mapping Crime: Principle and Practice* (National Institute of Justice, Washington, DC) (1999).
- [4] Rich, T., Shively, M., Abt Associates, I. and of America, U. S.: A methodology for evaluating geographic profiling software: final report, *Abt Associates Inc., Cambridge* (2004).
- [5] Snook, B., Taylor, P. J. and Bennell, C.: Geographic profiling: The fast, frugal, and accurate way, *Applied Cognitive Psychology*, Vol. 18, No. 1, pp. 105–121 (2004).
- [6] White, J. J. and Roth, R. E.: TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information, *Proceedings of GI-Science*, Vol. 2010 (2010).
- [7] Wang, X., Gerber, M. S. and Brown, D. E.: Automatic crime prediction using events extracted from twitter posts, *Social Computing, Behavioral-Cultural Modeling and Prediction*, Springer, pp. 231–238 (2012).
- [8] 桂一落合, 大祐鳥居, 悠 菊地: 位置に関連するツイート解析技術とその応用, *NTT DoCoMo テクニカル・ジャーナル*, Vol. 22, No. 2, pp. 30–35 (2014).
- [9] Fuchi, T. and Takagi, S.: Japanese morphological analyzer using word co-occurrence: JTAG, *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, Association for Computational Linguistics, pp. 409–413 (1998).
- [10] Dilrukshi, I., De Zoysa, K. and Caldera, A.: Twitter news classification using SVM, *Computer Science & Education (ICCSE), 2013 8th International Conference on*, IEEE, pp. 287–291 (2013).
- [11] Hofmann, T.: Probabilistic latent semantic analysis, *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., pp. 289–296 (1999).