

SNSの投稿内容を用いた居住地推定による アカウント到達可能性算出モデルの検討

吉國 綺乃[†]渡辺 知恵美[‡]小林 一郎[†][†]お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース[‡]筑波大学システム情報系 情報工学域

1 はじめに

近年、ソーシャルネットワーキングサービス（以下 SNS とする）の普及により、世界中で利用者が増えている。しかし一方では SNS でのトラブルが原因となり、利用者の個人情報取得されるという事例が発生している。攻撃者は、利用者が利用している SNS の情報をもとに個人情報を多く取得しようとする。これらの個人情報は自身が気が付かないうちに、投稿内容やプロフィールに自身で公開していることが多い。

SNS の利用において個人情報が取得することを防ぐためにも、自身が投稿した内容やプロフィールがどの程度のプライバシーリスクになっているかを把握する必要性が増している。我々は SNS におけるプライバシーリスクの提示指標として、アカウント到達可能性を定義している [1]。本論文ではアカウント到達可能性を求める具体的な手法のひとつとして、SNS の投稿内容に含まれる地域情報をもとにアカウント到達可能性算出モデルを検討する。文献 [2] では、Twitter の投稿内容のみを用いて利用者の居住地を推定するモデルを提案している。我々はこのモデルを参考にし、Twitter の投稿内容のみから利用者の居住地を推定し、別のアカウントを見つけ出すことを考える。

2 アカウント到達可能性

2.1 アカウント到達可能性算出式

アカウント到達可能性 (Account Reachability) とは、攻撃者が利用者の既知のアカウントから別のアカウントを見つけ出す可能性を表す。

A Study on a Model for calculating Account Reachability based on estimating User's Area of Activity with SNS Postings

[†]Ayano YOSHIKUNI(yoshikuni.ayano@is.ocha.ac.jp),

[‡]Chiemi WATANABE(chiemi@cs.tsukuba.ac.jp),

[†]Ichiro KOBAYASHI(koba@is.ocha.ac.jp)

[†]Advanced Sciences, Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Ohtsuka Bunkyo-ku Tokyo 112-8610

[‡]Faculty of Engineering, Information and Systems, Tsukuba University 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-0006, Japan

アカウント s_1 から別のアカウント s_2 を見つけ出す可能性は以下の式で求められる。

$$AR(s_1 \rightarrow s_2) = \max_{q \in Q} (AR(s_1, s_2, q))$$

$$Q = GenQueries(s_1.prof, s_1.msg).$$

$$AR(s_1 \rightarrow s_2, q) = Match(s_2, Cand(q)) * \frac{Score(s_1, s_2)}{\sum_{c \in Cand(q)} Score(s_1, c)}$$

$$Match(s_2) = \begin{cases} 1 & \text{if } s_2 \in Cand(q) \\ 0 & \text{else} \end{cases}$$

$GenQueries$ は s_1 のプロフィールや投稿内容から、 s_2 のアカウントを見つげ出すためのクエリを生成する式である。 $Score$ は s_1 と候補アカウント c との類似度を表す。これらの関数は、攻撃者が入手できるデータや利用できる技術に基づいて実装される。先行研究では技術知識を持たない攻撃者でもできる最もシンプルな方法として、 $GenQuery$ では Profile に含まれる名前、所属をキーワードに検索エンジンで検索する実装を、 $Score$ では検索エンジンのランキングの逆数をとる類似度計算を採用した。

2.2 アカウント到達可能性算出モデルの検討

先行研究ではプロフィールデータのみを用いたもっともシンプルな攻撃モデルを採用したが、データ分析技術を用いることで、ユーザの投稿履歴からも特徴的なキーワードを抽出し攻撃に利用することができる。本論文ではその 1 手法として、SNS の投稿内容に含まれる地域情報をもとにした攻撃モデルを求め、それを用いてアカウント到達性を定義し、検証を行う。まず 2 節にて述べた二つの関数の実装を以下のようにする。

- $GenQueries(s_1.prof, s_1.msg)$:

Tweet から地域名を取得しキーワードとする。検索エンジンを用いて別アカウントの候補を見つげ出す。

- $Score(s_1, c)$

検索エンジンでの検索結果のランキング結果をもとに類似度を計算する。

3 実験

3.1 実験設定

我々は Tweet に含まれる地域名が本当に別アカウントを見つげるキーワードになり得るのか検証するため

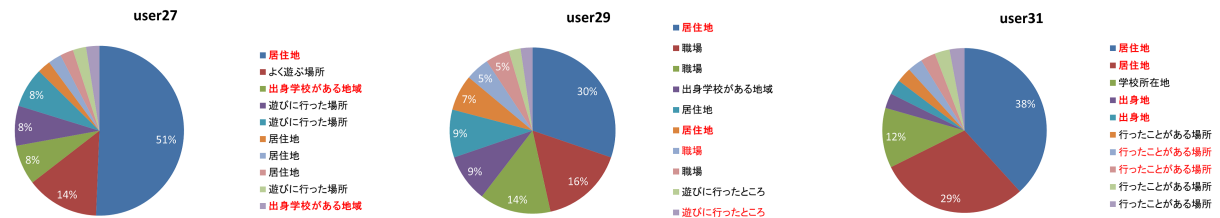


図 1: タイプ 3 の地域名分類

に予備実験を行った。本予備実験では、被験者を以下の 4 タイプに分類する。

1. 複数の SNS アカウントが同一人物と知られてよく、投稿内容に含まれる地域名も類似している
2. 複数の SNS アカウントが同一人物と知られてよいが、投稿内容に含まれる地域名は類似していない
3. 複数の SNS アカウントが同一人物と知られたくないが、投稿内容に含まれる地域名が類似している
4. 複数の SNS アカウントが同一人物と知られたくなく、投稿内容もきちんと使い分けができていない

我々が注目すべき利用者はタイプ 3 の利用者であり、この利用者は使い分けをしているつもりでできていない利用者である。

被験者は Twitter と Facebook の両方のアカウントを持つ 50 名であり、26 名は両方のアカウントが第三者に同一人物だと知られても良いと考えている被験者、24 名は同一人物だと知られたくない被験者である。各被験者に対し、tweet と Facebook の投稿内容また Facebook のプロフィールから地域名を取得し、取得した地域名を目視でどれだけ類似しているか比較する。

Tweet 数は各被験者 1000 件、Facebook の投稿数は約 100 件収集した。取得した文章に対しストップワードの除去を行い、形態素解析器 MeCab を用いて形態素解析を行った。

3.2 実験結果

被験者のうち 28% がタイプ 3 であることが分かった。タイプ 3 に分類された被験者を対象に Tweet から抽出した地域名上位 10 件が被験者にとってどのような地域なのか、またどのような地域名が Facebook から抽出した地域名と類似しているのか検証を行った。タイプ 3 の被験者から代表 3 名の結果を図 1 に示す。

赤い太文字で書かれている地域は、Facebook から抽出した地域と同一の地域である。彼らの特徴として、居住地が Tweet 内に出現する地域名の中で最も出現しており、またそれらが Facebook 上でも取得可能であることがいえる。被験者にヒアリングを行ったところ、友人とのやり取りなどを通していつの間にか地域名をつぶやいてしまっているとの回答を得た。居住地は SNS 利用者にとっても漏えいしては困る個人情報の

ひとつであり、地域名を利用してアカウント到達可能性を求める手法は有用であるといえる。

しかしこれらの被験者に対しアカウント到達可能性を求めたところ、全員 0% であった。これは地域名だけでは多くの利用者が該当するため、個人を特定しづらいことを示す結果である。

4 まとめと今後の課題

近年 SNS の普及とともに、SNS の利用において自身が投稿した内容やプロフィールがどの程度のプライバシーリスクになっているかを把握する必要性が増している。我々は、SNS におけるプライバシーリスクの提示指標として、アカウント到達可能性を定義した。本論文ではアカウント到達可能性を算出するモデルの一検討として、居住地推測を用いたモデルの検討を行った。

Twitter から地域名を取得し、Facebook の投稿内容やプロフィールから抽出できる地域名との類似度をはかることでアカウント到達可能性を求めることを目標とする。本論文では、Twitter から抽出した地域名が実際に利用者の個人情報に近いものであるのか、また Facebook から抽出できる地域名との類似性はあるかを検証するため予備実験を行った。

実験の結果、Twitter から抽出した地域名は実際に利用者の個人情報に近いものが多く抽出され、また Facebook から抽出できる地域名との類似性もみられた。

我々の目標は、利用者自身が SNS の利用において自身が投稿した内容やプロフィールがどの程度のプライバシーリスクになっているかを把握することである。本論文では地域情報のみを用いたアカウント到達可能性算出モデルの検討を行ったが、アカウント到達可能性を算出する手法はさまざまある。今後これらの検討も行っていきたい。

参考文献

- [1] Ayano YOSHIKUNI, and Chiemi WATANABE "Account Reachability : A Measure of Privacy Risk for Exposure of a User's Multiple SNS Accounts", iiWAS2013
- [2] Zhiyuan Cheng, James Caverlee, and Kyumin Lee "You Are Where You Tweet: A Content-Based Approach to Geolocating Twitter Users", CIKM '10