1ZA-02

ロゴマークを利用したフィッシングサイト検知手法の提案

高木　秀輔†　寺澤　卓也†
東京工科大学メディア学部†

1. はじめに

近年、Web を舞台としたフィッシングという詐欺の手法が、これまで以上に巧妙化している。フィッシングとは、金融機関などの正規の Web サイトを騙り、消費者から個人情報を搾取するものである。

本研究では、フィッシングサイトが正規のサイトの模倣であること、正規のサイトには多くの場合、ロゴマークが記載されているといった点に着目し、フィッシングサイトを検知する手法を提案する。また、これまでの研究の手法と比較することで、本手法の有用性を示す。

2. 関連研究

加藤らは、コンテンツベースの検知手法の大規模な評価を行った[1]。フィッシングサイトと正規サイトの内容は酷似しており、そこに出現する言葉や見た目には同じ特徴がみられる。コンテンツベース方式とは、このような類似性に着目することで、フィッシングサイトの検知を行うという方式のことである。

この研究により、日本語、英語のいずれにおいても、高い検知性能が示され、コンテンツベース方式の有効性が確認された。

3. ロゴマーク抽出による検知手法

正規サイトには、各サービスや企業のロゴマークが記載されている。フィッシングサイトは、基本的に正規のサイトを模倣して作っているため、正規サイトに記載してある正規のロゴマークは、フィッシングサイトにも記載してある場合が多いと考えられる。また、フィッシングサイトの特徴である、存続期間が短いということと他のサイトからリンク付けされないということから、フィッシングサイトは通常の検索サイトでは、検索結果の上位には示されない。

よって、現在閲覧しているサイトからロゴマークを抽出し、画像検索を行うことで正規のサイトを見つけることができる。そして、検索結果上位の正規サイト候補と URL の比較を行うことで、フィッシングサイトかどうかの判定を行う。これらを実装するために以下の 4 つの手順が必要である。

① ロゴマーク抽出のための機械学習

Web ページからのロゴマーク抽出には、画像検出アルゴリズムの YOLO[2]を利用する。

画像検出に必要な学習のために、Web ページの中のどの部分がロゴマークなのかを示すデータセットを約 700 件作成した。

② ロゴマーク抽出

閲覧している Web ページの全体のスクリーンショットを撮影し、学習データを基に物体検出を行い、ロゴマークを検出する。

抽出したロゴマークを使用して画像検索を行い、結果上位の URL を取得する。検索結果の取得には、本来であれば、API 等を利用することが好ましいが、主要検索エンジンの多くは画像検索の API を提供しておらず、API を利用できる検索エンジンでも芳しい結果を得られなかったため、Selenium を利用したスクレイピングという手段を採用した。

③ URL 比較

閲覧しているサイトと、取得した URL の比較を目視にて行い、フィッシングサイトかどうかの判定を行う。

次の図 1 はこれらを図式化したものである。

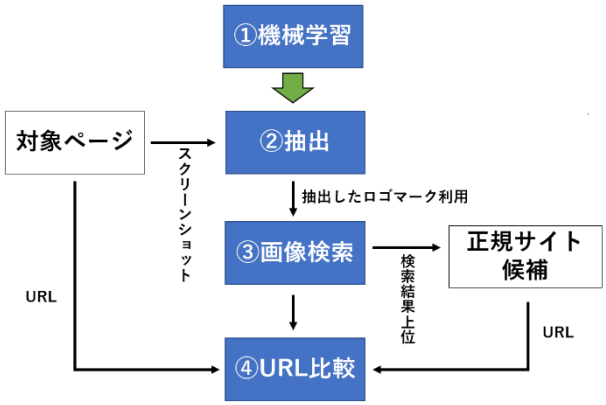

図 1：手法の概要

A novel phishing site detection method using logos on the websites.
†Shusuke Takagi, Takuya Terasawa
(Tokyo University of Technology School of Media Science)

4. 評価

4.1. 評価方法

評価には、フィッシング対策協議会[3]に掲載されていた 10 のフィッシングの事例から、それぞれの正規サイト、フィッシングサイトを用い、以下の 2 点について行う。

- フィッシングサイト検知率

本研究では、フィッシングサイトを、正規サイトを導出したうえでフィッシングサイトと正しく判定した率を示す。

- 正規サイト導出率

本研究では、正規サイトを正規サイトと正しく判定した率を示す。

4.2. 既存手法との比較方法

本来であれば、実際に先行研究で使用された手法を再現し、同じ環境下で比較することが望ましいが、次の理由によりそれは難しいと考える。

- 既存手法を再現可能であるかが不明。
- 先行研究で使用されたサイトの URL が不明。
- 使用されたフィッシングサイトは存続していない。

そのため、上述した 2 点の比較を既存手法との比較とする。これらの評価方法は、前述した加藤らの研究で使われており、既存手法との比較に適していると考えられる。

4.3. フィッシングサイト検知率

今回実験を行った 10 件のうち、意図したとおりにフィッシングサイトと判定できたものは、4 件であった。1 件が適切なロゴマークを抽出することができなかったこと、5 件が不適切なキーワードが付与されてしまったことが正規サイトを導出することできなかった理由である。1 点目に関しては、ロゴマークが複雑なデザインになっており、ロゴマーク全体をロゴマークと判別することができず、一部分のみをロゴマークとして抽出してしまったことが原因であると考えられる。2 点目に関しては、画像検索を行う際に自動的に付与される検索キーワードが、意図しているものとは全く違うものが選定されてしまうことが原因である（図 2）。

図 2：PayPay[4]のロゴマークの検索結果

有名な企業やサービスであれば、ロゴマークと関連した語句がキーワードに選定され適切な結果を得ることができる場合が多いが、それ以外のものは、ロゴマークに多く使われている色をキーワードに選定している場合が多かった。

4.4. 正規サイト導出率

今回実験を行った 10 件のうち、正規サイトを導出できたのは 4 件であった。また、1 件が適切なロゴマークを抽出することができなかったこと、5 件が不適切な検索キーワードが付与されてしまったことが導出できなかった原因といえる。

4.5. 既存手法との比較

上述した 2 点の評価方法に関して先行研究と比較を行うと、いずれも本研究の方が低い結果となった。これは、上述した不適切な検索キーワードの付与が主な原因であると考えられる。

5. まとめ

本論文では、様々な種類のフィッシングサイトに対応できる手法の提案、また、既存手法との比較により、本手法の有用性を示すことを目的とし、ロゴマークに着目した検知、評価を行った。先行研究と比較すると低い精度となったが、機械学習を利用することにより、ほとんどの場合、適切なロゴマークを抽出し、検知の足掛かりとすることができた。そのため本手法の可能性は示されたといえる。

今後の課題として、正規サイトを見つけることができないという問題があったため、検索方法を再検討し、より精度の高い機械学習を行う必要がある。また、テンプレートマッチングのような手法を利用し、限られたデータの中からであれば正確な判定を下すことができるかといった実験を行う予定である。

参考文献

[1] 加藤慧, 小宮山功一朗, 瀬古敏智, 一瀬友祐, 川野耕平, 吉浦裕 “コンテンツベースフィッシングサイト検知手法の大規模実例評価と改良” Vol.2010-DPS-142 No.44, 2010 年 3 月 5 日

[2] “YOLO:Real-Time Object Detection”, https://pjreddie.com/darknet/yolo/ (参照 2019-7-28)

[3] “フィッシング対策協議会　Council of Anti-Phishing Japan”, https://www.antiphishing.jp/(参照 2019-8-5)

[4] “PayPay - QR コード・バーコードで支払うスマホ決済アプリ”,https://paypay.ne.jp/(参照 2020-1-9)