パターン認識を用いたフィッシングサイトの検知手法の提案

Proposal for the Detection Method of a Phishing Sites by Using Pattern Recognition

1541038　笠原 拓也
Takuya KASAHARA
指導教員　秋葉 知昭

In this study, I proposed algorithms for the image recognition and natural language processes using deep learning. Then, I performed a detection test of the phishing site with the learned model.

1. 緒　言

近年, 再び, オンラインショップやオンラインサービスのサイトを装い，アカウント情報やクレジットカード番号などの個人情報を詐取するフィッシング詐欺が問題となっている．個人や企業などの組織がインターネットやコンピュータを安全に使い続けられるように，ウィルスやフィッシング詐欺により個人情報や営業機密などの情報が外部に漏れたり，悪用されたりしないようにセキュリティ対策[1]をする必要がある．最近では，パソコンだけでなく，スマートフォンも同様に電子メールからフィッシングサイトに誘導される手口[4]が増えてきている．電子メールの送信者名を偽装し，緊急を装う文面だけでなく，正規のサイトと区別がつかない偽のサイト(フィッシングサイト)に偽造するなど，フィッシングサイトの巧妙化が進んでいるため，既存の対策方法では，Web サイトの正当性を判断するのは困難である．

そこで本論文では，フィッシングサイトの検知手法として, ディープラーニングによる、画像認識と自然言語処理のアルゴリズムを作成し、学習と評価を行うことで，フィッシングサイトをリアルタイムで検知する手法を提案する．

2. 検知手法の提案

2.1 検知手法の概要

本研究は，画像と HTML ソースの 2 つの観点からディープラーニング(以下 DL)で特徴を学習させ, フィッシングサイトか否か判別させる手法である．判別する前に，画像と HTML ソースそれぞれで学習アルゴリズムを作成する．判別する際は, ディープラーニングツールである Neural Network Console（以下 NNC）を使用する．

2.2 収集データ

画像と HTML ソースの判別を行うためには，まずフィッシングサイトと正規サイトのデータを収集する必要がある．収集データは, サイトのキャプチャ画像，HTML ソースの 2 種類である．フィッシングサイト 200 件,正規サイト 200 件からそれぞれ画像 200 件，HTML ソース 200 件のデータを用いる．

2.3 画像認識学習アルゴリズム

NNC で構築した画像認識[2][3]の学習アルゴリズムのイメージを図 1 に示す．入力層として Input，畳み込みを行う Convolution，画像の縮小を行う Max Pooling，全結合を行う Affine，数値の活性化を行う PReLU，Tanh，Softmax，出力層として Categorical Cross Entropy，Batch Normalization を用いた．Convolution で画像の特徴を抽出し,Max Pooling で画像の特徴を残しつつ縮小を行う．そして,抽出と縮小を繰り返し, 1 次元配列に変換され，最後に確率に変換後結果を出力する．この Input から Categorical Cross Entropy までの処理を正解との差が小さくなるようデータの数だけ繰り返し学習を行う．関数だけでなく, 特徴マップの生成数も精度に関わった．生成数を徐々に増やすことで, 学習も徐々に行うことができ，精度向上につながったのではないかと考える．

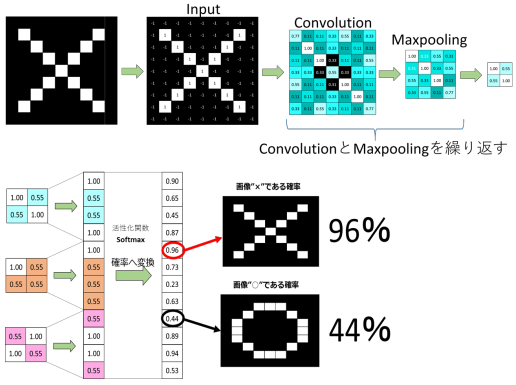

図 1　画像認識学習アルゴリズムの仕組み

千葉工業大学　社会システム科学部　経営情報科学科
2018 年度卒業研究概要

2.4 自然言語処理学習アルゴリズム

NNC で構築した自然言語処理[5]の学習アルゴリズムのイメージを図 2 に示す．入力層として Input，全結合を行う Affine，数値の活性化を行う Tanh, Sigmoid, 出力層として Binary Cross Entropy を用いた．数値化し，正規化を行ったソースに重み付けがされ，1 次元の配列に変換される．変換されたソースを活性化関数で-1 から 1 の間に補正し，再び 1 次元配列に変換し，最後に確率に変換して結果を出力する．この Input から Binary Cross Entropy までの処理を正解との差が小さくなるようデータの数だけ繰り返し学習を行った．

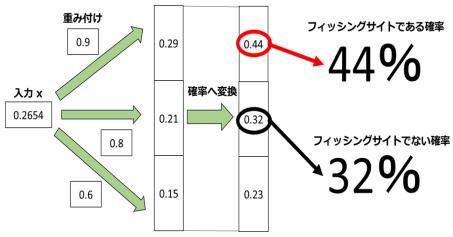

図 2　自然言語処理学習アルゴリズムの仕組み

3. 結果及び考察

3.1 画像認識

画像認識には，学習に用いたデータとは異なるデータを用いて評価した．本研究において NNC で作成した学習アルゴリズムの結果は，正検出 90%，誤検出 10%と高い精度を示した．これは，学習用とは異なるデータ，未知のデータを用いた結果であることから，汎用性があることが判った．正検出，誤検出それぞれについて，サイトの傾向を分析したところ，正検出は，比較的複数枚の画像から構成されていたサイトが多数であった．サイトのロゴや広告画像，イメージ画像等などから特徴となる部分を抽出することができたものと考える．反対に誤検出では，画像が使われずに背景が白や黒一色，ロゴやその他の画像が配置されていないものがほとんどであった．恐らくこれは，特徴として抽出することができる画像が配置されていない，画像が少ないなどの原因が考えられる．

3.2 自然言語処理

自然言語処理も同様に，学習に用いたデータとは異なるデータを評価に用いている．NNC で作成した学習アルゴリズムの結果は，正検出 68.14%，誤検出 31.86%の精度を示した．この精度では，十分とは言えない．更なる改良が必要である．正検出，誤検出についてサイトの傾向を分析したが，正誤どちらも規則性を見つけることができなかった．また，同じ文字列が正誤両方に分類されていた例がいくつも確認された．恐らく HTML ソースの数値化に問題があるのではないかと考えられる．

4. 結　言

本研究では，パターン認識を用いたフィッシングサイト検知手法を提案し，DL の学習及び評価を行った．画像認識の結果は，フィッシングサイトと正規サイトを 90%の精度で判別することを確認した．この結果から，DL でパターン認識である画像認識を用いた検知は，フィッシングサイト判別手法として十分有効で，汎用性があると言える．自然言語処理の結果の場合は，フィッシングサイトと正規サイトを 68.14%の精度で判別することを確認した．しかし，正検出及び誤検出に規則性が見られなく，フィッシングサイトの判別手法として精度は不十分であり，改良が必要である．これらの結果から，DL でパターン認識を用いたフィッシングサイトの検知手法は有効であることが確認された．画像認識については，巧妙化が進むフィッシングサイトに対して高い検出精度が期待でき，人の視覚では判別することができない場合でも，従来のフィッシング詐欺対策とは違って正当性をリアルタイムで判断するので安心して任せることができる．

今後は画像及び HTML ソースのデータを吟味し，更新していくことで精度の向上及びフィッシング詐欺の犯人に対して先手が打つことができるのではないかと考えられる．

文　献

[1] 総務省：安心してインターネットを使うために,http://www.soumu.go.jp/main_sosiki/joho_tsusin/security/cmn/download/kokumin-security_enduser.pdf (2018 時点)
[2] MathWorks：画像認識とは,https://jp.mathworks.com/discovery/image-recognition.html (2018 時点)
[3] TECHACADEMY：画像認識とは,https://techacademy.jp/magazine/17068 (2018 時点)
[4] フィッシング対策協議会：フィッシングとは,https://www.antiphishing.jp/consumer/abt_phishing.html (2018 時点)
[5] Steven Bird, Ewan Klein, Edward Loper：入門 自然言語処理，オライリージャパン(2010)