自然言語処理の深層学習によるフィッシングサイト検知手法の改良

Proposal for Improvement Idea for the Phishing Site Detection Method by Deep Learning of the Natural Language Processing

1641046　小林 貴章
Takaaki KOBAYASHI
指導教員　秋葉 知昭

In this study, I take about natural language processes using deep learning. In this study, I proposed improvement idea of the natural language processing for the detection method. I showed proposed idea is better than previous study.

１．緒　言

近年, 銀行や郵便局など公的な機関や, 有名ブランドによるオンラインサービスのような信用のあるサイトを装い，クレジットカード番号や ID・パスワードといった個人情報を騙し取るフィッシング詐欺は，インターネットの普及に伴い年々増加している．最近では，掲示板などの SNS で直接やり取りを行いながらフィッシングサイトへ誘導するなど，ネットを使用する人々の行動に即した新しい手段を作り出している．こういった新しい手段を周知し警戒を呼び掛けたり，サイトの閉鎖やブラックリスト化を行うことによる対策は常に後手に回り，早期の被害を減らすことは難しい.

この問題に対して，笠原[1]は画像と HTML ソースのパターン認識を，ディープラーニングを用いて行い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した．この手法による対策はリアルタイムで検知を行うため，フィッシングサイトに対して先手を打ち，被害を防ぐことが期待できる. この結果, 画像認識では実用的な数値を得たが, 自然言語処理では実用に耐えうる精度を得ることができなかった.

本論文では，自然言語処理によるフィッシングサイト検知手法の精度を向上させることを目的とし，ディープラーニングの学習及び評価の結果について述べる.

２．検知手法の提案

2.1 検知手法の概要

本研究は，HTML ソースの特徴をディープラーニング(以下 DL)で学習させ，フィッシングサイトか否か判別させる手法の提案を目的としている．実際の判別には DL ツールである Neural Network Console[2](以下 NNC)を使用する.

2.2 収集データ

HTML ソースによる DL を行うためには，まずフィッシングサイトと正規サイトそれぞれの HTML ソースデータを収集する必要がある．本研究では先行研究との比較を行うため，笠原の先行研究[1]で使用されたデータを使用した．フィッシングサイト 200 件，正規サイト 200 件の収集データをそれぞれ学習用，評価用に用いる.

2.3 HTML ソースの数値化

笠原の先行研究[1]では NNC を用いて HTML ソース全文を学習させると学習時間が膨大になり，学習が難しくなるため URL のみを抽出し，その URL を形態素解析によって分割し, TF-IDF 法を用いて数値化した.

TF-IDF とは，文書内でのある単語の出現頻度を意味する TF(Term Frequency)と，ある単語が含まれる文書の割合の逆数 IDF(Inverse Document Frequency)を掛け合わせることで，単語の重要度を測る手法である．本研究では数値化の方法によって精度を改善するために，TF-IDF 法の代わりに okapiBM25[3]を使用した．TF-IDF では文書の総単語数による差が大きく，文書毎の総単語数にばらつきが多い URL の比較では偏りが大きく出てしまう欠点がある.

okapiBM25 はその問題を文書の総単語数を示す DL(Document Length)の用を加えることで文章による偏りを緩和した手法である．式(1)に okapiBM25 の計算式を示す. また, 場合に合わせてパラメータを調整する必要がある.k_1は単語の出現頻度から計算した重要度(TF-IDF 値)の影響の大きさを調整するパラメータであり，k_1=1.2 もしくは 2.0 が使用される．また b は主に文書の単語数による影響の大きさを調整するパラメータであり，0.0 から 1.0

千葉工業大学　社会システム科学部　経営情報科学科
2019 年度卒業研究概要

の間で設定する．本研究では，それぞれ最も効果的であることが確認されているk_1=2.0，b=0.75 を採用している．本研究では一般的に効果的であるとされているk_1=2.0，b=0.75 を使用している。NNC は入力する数値を 1.0 から−1.0 間に調整する必要があるため，数値変換後にデータの中で最も大きな絶対値を取り出し，全てのデータをその値で割ることによって正規化した．

$$\text{okapiBM25}(t_k,d_i) = \text{idf}(t_k)\cdot\frac{tf(t_k,d_i)\cdot k_1+1)}{tf(t_k,d_i)+k_1\cdot(1-b+b\cdot\frac{dl(d_i)}{avgdl})} \quad (1)$$

$$\text{tf}(t_i,d_i)=\frac{\text{文書}d_i\text{内の単語}t_i\text{の出現回数}}{\text{文書}d_i\text{のすべての単語の出現回数の和}} \quad (2)$$

$$\text{idf}(t_i)=\log\left(\frac{\text{層単語数}-\text{単語}t_i\text{が出現する文書数}+0.5}{\text{単語}t_i\text{が出現する文書数}+0.5}\right) \quad (3)$$

$$\text{dl}=\text{文書}d_j\text{のすべての単語の出現回数の和} \quad (4)$$

$$\text{avgdl}=\text{文書全体の平均DL値}=\frac{\sum_{d_i\in D}dl(d_i)}{|D|} \quad (5)$$

2.4 自然言語処理学習アルゴリズム

NNC 上で行う自然言語処理学習アルゴリズムのイメージを図 1 に示す．入力層に Input，全結合を行う層 Affine, 数値の活性化を行う層 Tanh, Sigmoid, 出力層に 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−1.0 から 1.0 間に補正し，再び 1 次元配列に変換し, 確率に変換して結果を出力する．この一連の処理をデータの数だけ繰り返す学習を行った．

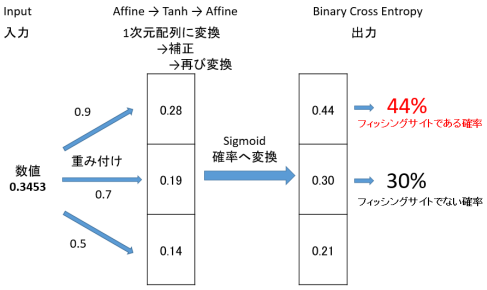

図 1　自然言語処理学習アルゴリズムの構造

3．結果及び考察

NNC を用いて自然言語処理の学習及び評価を行った結果, 77.25%の正検出が得られ, 特徴マップ生成数や層の増加による数値の改善には規則性が見られた．しかし誤検出の規則性が見られない点から，信頼性に疑問が残る．また精度は改善されたが，フィッシングサイト検知手法として十分な実用性がある数値とは言えず，更なる改良が必要である．今後の改善点として，URL の抽出や数値化もしくはその後の段階で, 学習のための調整を行い, 結果が変化するかの検証が必要だと考えた．

4．結　言

本報告では，okapiBM25 を用いた自然言語処理フィッシングサイト検知手法の改良を提案し，DL の学習及び評価を行った．自然言語処理による結果から 77.25%となり先行研究[1]を上回る数値を得られたが，精度 77.25%という数値でもフィッシングサイト判別手法として十分な数値とは言えず，実用性があるとはいえない．しかし改善の余地は見られるため，正規化の前段階での数値の調整や，抽出する URL データの偏りを解消することが必要だと考えられる．

これらの結果から，okapiBM25 をはじめとした手段による自然言語処理の DL の精度向上が確認された．しかし検知手法として実用化するためには，前述のようにより別の方向性から改善を目指すか，あるいは根本的に手法を変えることによって十分な精度を得る必要があると考えられる．

画像認識によるフィッシングサイトの検知手法は既に有効な結果を出しているが，画像認識処理だけでなく，自然言語処理によるフィッシング対策システムを併用することで，より高度な対策効果が期待できる．手口の巧妙化および件数の増加が進むフィッシングサイトに対して，パターン認識による DL を用いた検知手法は，被害を未然に防ぎ，頻繁な更新を必要としない点が従来のフィッシングサイト対策に比べて優れていると言える．今後は更新によって精度を高めていくことで，人力による対策の手間を減らし，フィッシングサイトの早期発見・対策が可能になると考えられる．

文　献

[1] 笠原拓也 : 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，2018 年度千葉工業大学卒業研究 (2019)

[2] SONY : Neural Network Console，http://dl.sony.com/ja/ (2020/1/23 時点)

[3] MIERUCA :【技術解説】単語の重要度を測る？TF-IDF と Okapi BM25 の計算方法とは，https://mieruca-ai.com/ai/tf-idf_okapi-bm25/(2020/1/23 時点)