6ZC-03

機械学習を用いた悪意のあるURLの検出の一手法

Zhang Weiming† 高田豊雄†
岩手県立大学ソフトウェア情報学研究科†

1 はじめに

インターネットの普及に伴い，その影響は我々の生活の各面に深く浸透している．ユーザーはURLを介してインターネット上の多種多様な情報に直接的あるいは間接的にアクセスする．しかし，不正なウェブサイトや金融詐欺といった危険も潜んでおり，URLには良質なものと悪質なものが混在している．フィッシング，トロイの木馬，マルウェアなど，多様な攻撃が悪意のあるURLを通じて実施されている．このような問題に対処するため，悪意のあるURLを検出し，ユーザーをこれらの脅威から保護するためのセキュリティ評価技術や手法の開発と研究が不可欠である．

2 関連研究

Shengら[1]はシグネチャに基づく悪意のあるURLの検出方法を提案した．新しいURLにアクセスするたびに，新しいURLにアクセスするたびに，悪意のあるURLを蓄積したデータベース(ブラックリスト)のクエリが実行される．もしそのURLがブラックリストに載っていれば．悪質と判断され，警告が表示される．

この方法の主な欠点は，与えられたリストにない新しい悪意のあるURLを検出することが非常に困難であり，検出率と適時性も低い．一方，Shenら[2]は悪意のあるURL検出の研究で，モデルに大きく貢献する最適な特徴部分集合を選択するという特徴選択方法を設計した．各特徴がランダムフォレストの各木にどれだけ貢献しているかを判断し，その平均値を取り，最終的に特徴間の貢献を比較する．この方法により，必要な特徴の選択と無関係な特徴を削減することが可能となる．しかし，特定の特徴の貢献傾向に過度に依存すると，繊細な特徴を見逃す可能性がある．また、特徴の貢献度は時間とともに変化する可能性があるため定期的にモデルを更新する必要がある．

3 提案手法

本稿では効率の良い特徴抽出手法を新たに提案する．調査論文[2]によると，ランダムフォレストが性能が一番よいアルゴリズムである．本研究では，それら既存の機械学習研究を踏まえ，TF-IDFアルゴリズムによるURLのテキスト特徴の抽出とランダムフォレストに代わるモデルとして深い森(multi-Grained Cascade Forest, 以下gcForest)を使用した悪意のあるURLの検出に取り組む．

TF-IDFは，テキストデータから特徴を自動的に抽出する．全データセットに基づいて特徴を計算するため，新しいパターンに自然に適応する．単語レベルで動作するため，より細かい粒度の特徴表現を提供することができる．また，大規模なデータを高い計算効率での処理に適している．

gcForestは，様々なタイプの特徴を持つデータを処理するのに特に適した機械学習モデルであり，2017年にZhouら[3]により提案された．gcForestはカスケード構造を採用している．これは多層構造で，各層には複数のランダムフォレストがあり，本稿で採用するgcForestでは各層は4つのランダムフォレストから構成されている．2つのランダムフォレストと2つのエクストリームフォレストがあり，各フォレストはデータに対して訓練を行い，入力されたデータを分類した結果が出る．この結果はフォレストが生成するクラスベクトルと呼ばれる．多くの研究において，gcForestモデルは，分類問題における顕著な効果により注目されている．腫瘍分類[4], Eコマース商品分類[5]まで, gcForestはさまざまなデータセットで有効な分類能力を発揮している．これらの先行する成功した応用例に基づき，悪意のあるURLの分類においてもgcForestモデルが高精度な識別を実現できると考えられる．

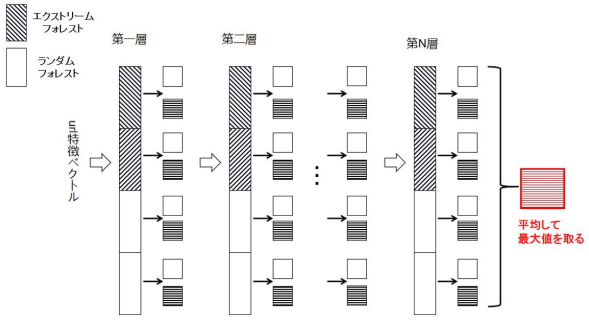

図 1 gcForestによる悪意のあるURL検出の流れ

A method of detecting malicious URLs using machine learning
† Weiming Zhang and Toyoo Takata, Graduate School of Software and Information Science,Iwate Prefectural University

4 結果

まず, 悪意のある URL と良性 URL のデータセットを収集する. Patgiri ら[6]によれば, 80:20 の割合で分割する方が, より精度の高い分類が可能となることが示された. 良性 URL と悪意のある URL 合計 424,000 件の URL を取得した. さらに 424,000 件の URL をランダムに並べ替え, 340,000 件を訓練用, 84,000 件をテスト用に分割した. 次に, このデータセットは前処理を行い, 機械学習モデル処理できる形式に変換した後, サンプル中の各単語の TF-IDF 値を計算し, それらの特徴ベクトルを抽出し, それらの特徴ベクトルを使用して gcForest モデルを訓練する. 訓練されたモデルの性能は再現率, 正解率, 適合率と F1 スコアの指標を用いて評価される.

まず, 先行研究において提案された特徴貢献傾向に基づく特徴選択方法と, 本研究で提案する TF-IDF を用いた特徴抽出方法の比較実験を行った. 本実験ではモデルの選択については, 先行研究で最も精度が高かったランダムフォレストモデルを用いる. 実験結果を図 2 に示す.

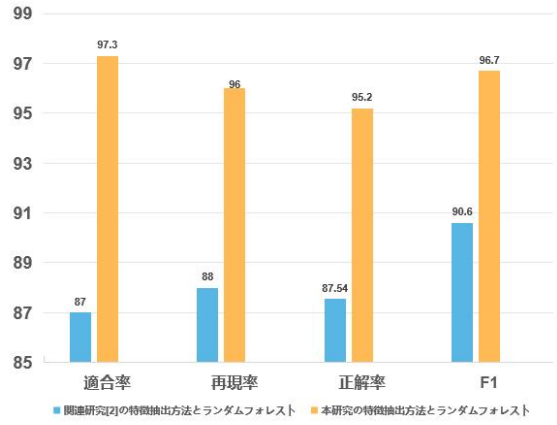

図 2 実験結果 (1)

この結果, TF-IDF 特徴抽出方法より精度が高いでことが分かった.

次の実験では同じ特徴抽出方法—TF-IDF を用いる. モデルについて本稿で提案された gcForest とランダムフォレストの比較実験を行う. 実験結果を図 3 に示す.

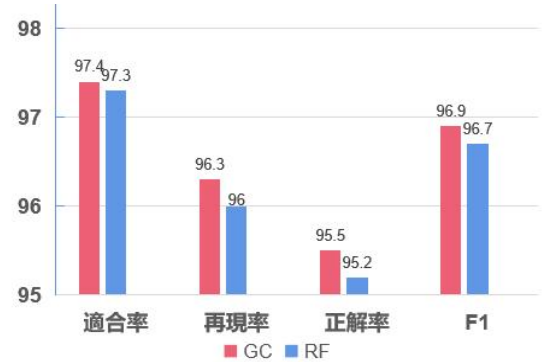

図 3 実験結果 (2)

この結果によると, gcForest は, 正解率, 再現率, 適合率, F1 スコアの各指標において, 従来手法であるランダムフォレストより高くなっている.

5 まとめ

本稿では, 悪意のある URL の検出に関する研究を行った. 特徴抽出には TF-IDF を利用し, 実験結果から, TF-IDF がより優れた分類効果を示すことがわかった. また, 悪意のある URL の識別には gcForest を採用し, 高い精度が得られることを示した.

現在流行している深層学習アルゴリズムを使用していないため, 今後は, CNN や LSTM などのより多様なモデルを悪意のある URL の識別に応用することで, より高い精度を得る可能性を検討する.

参考文献

[1] S. Sheng, B.Wardman, G.Warner, L.F. Cranor, J. Hong, and C. Zhang, An empirical analysis of phishing blacklists, Proceedings of Sixth Conference on Email and Anti-Spam (CEAS), 2009.

[2] S. He, J. Xin, H. Peng, and E. Zhang, Research on Malicious url Detection Based on Feature Contribution Tendency, IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2021.

[3] Z. Zhou, and J. Feng, Deep Forest: Towards An Alternative to Deep Neural Networks, IJCAI, pp. 3553–3559, 2019.

[4] Z. Chen, X. Sun, and L. Shen, An effective tumor classification with deep forest and self-training, IEEE Access,vol.9, pp.100944–100950, 2021.

[5] J. Dai, T. Wang, and S. Wang, A deep forest method for classifying e-commerce products by using title information, ICNC 2020, Big Island, HI, USA, February 17-20, 2020, pp. 1–5, IEEE, 2020.

[6] R. Patgiri, H. Katari, R. Kumar, and D. Sharma, Empirical Study on Malicious url Detection Using Machine Learning, Distrib. Comput. Internet Technol., Springer, pp. 380-388, 2019.