講演番号：84

機械学習を用いた良性・悪性 URL 分類法の検討

A－7　Study of classification method of clean/malicious URL using machine learning

小出 遼† 笠間 貴弘‡ 宮保 憲治†
Ryo KOIDE† Takahiro KASAMA ‡ Noriharu MIYAHO†
† 東京電機大学 情報環境学部 情報環境学科 ‡ 国立研究開発法人情報通信研究機構
† School of Information Environment. Tokyo Denki University ‡ National Institute of Information and Communications Technology

1. はじめに

近年多種多様なマルウェアが蔓延し、その被害は増加している。これらのマルウェアは、攻撃サイトと呼ばれるマルウェア配布元サイトからユーザーの端末にダウンロードされ、個人情報収集などの活動を行う。その中でも、ExploitKit と呼ばれる攻撃ツールを使用し、攻撃者が簡単にかつ悪性サイトを多数作成する状況が報告されている。

本検討では、ExploitKit を使用した可能性が高い悪性サイトは URL が類似していることに着目し、悪性サイトの URL を収集した上で良性・悪性サイトを SVM と Random Forest アルゴリズムを用いて分類を行う機械学習システムを構築した。本論文では当該システムの性能評価の一環として、分類精度について評価した結果を報告する。

2. 提案手法

2.1 データ収集の概要

本検討では悪性サイトの URL 収集にはマルウェアの配布元サイトのデータベースである MalwareDomainList[1]を使用した。良性サイトのデータ収集には Python のフレームワークの一つである Scrapy を利用した低対話型クローラを用いて、キュレーションサイト[2]の一つである、はてなブックマーク[3]のクローリングを行い URL データを収集した。

2.2 機械学習システム

本検討で使用した機械学習システムの性能評価を行うため、2.1 で収集したデータを、文書中に出現した単語の重要度を評価する手法である TF-IDF(索引語頻度逆文書頻度)[4]を用いてベクトル化した。単語分割は URL に含まれるドットやスラッシュ単位で行った。さらに、そのデータに正解ラベル(良性: 0, 悪性: 1)を付加した教師データを作成した。このデータは 2.3 で行う性能評価実験で使用する。提案する処理フローを図 1 に示す。

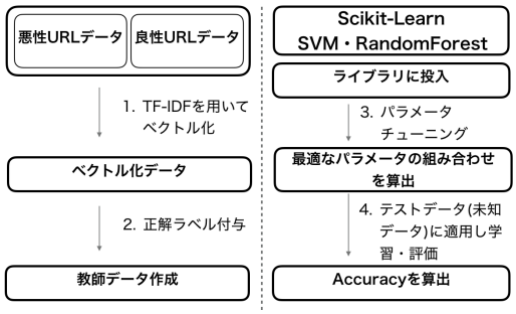

図 1 提案する処理フロー

2.3 性能評価

2.2 で作成した教師データを用いて、性能評価実験を行った。本検討では教師データのモデルの分類精度を検証するために、予めグリッドサーチを併用した交差検証を利用して、Gamma・C 等のパラメータチューニングを行った。グリッドサーチのパラメータ用に SVM は Gamma と C(それぞれの範囲: 0.001、0.01、0.1、1、10、100)、Random Forest は max_depth(決定木の最大深さ、範囲: 2、3、4、5、6、7)と min_samples_leaf(決定木の葉に分類されるサンプル数、範囲: 1、3、5、7、10)を使用した。本検討ではモデル分類時の過学習を防ぐため、交差検証の手法に 5 分割交差検証を用いた。性能評価実験結果を表 1 に示す。実験には良性サイト・悪性サイト共に 500 件、合計 1000 件使用した。

表 1 性能評価実験結果

	Accuracy(テストデータに対する正答率)	最も評価精度が良いモデルのパラメータの値
SVM	96%	C: 10, Gamma : 0.1
RandomForest	92%	max_depth: 7, min_samples_leaf: 1

3. 実験結果

表 1 より、URL による分類を用いた場合、 SVM を使用することによりテストデータに対して Random Forest より高い正答率(96%)が得られることを確認した。本検討のように比較的少ないデータ数である場合の URL 分類には SVM がより適切であることを確認した。本検討ではテストデータは未知データとして扱っており、96%の正答率であるため、当該システムはアルゴリズムに SVM を用いることによりモデル分類に対して Random Forest より高い分類精度を発揮すると考えられる。

4. まとめ・今後の展望

本稿では、良性サイト・悪性サイトを、機械学習を用いて URL により分類する手法を提案した。今後は教師データ数のさらなる拡充を行い、他の機械学習システムを使用した場合の性能評価実験も行う予定である。

5. 参考文献

[1] https://www.malwaredomainlist.com/
[2] https://www.weblio.jp/content/キュレーションサイト
[3] https://b.hatena.ne.jp/
[4] J. Ramos. Using tf-idf to determine word relevance in document queries. In ICML, 2003.