Computer Security Symposium 2020
26 - 29 October 2020

ブラウジング履歴情報に基づく悪性サイトの事前検知手法の改善

巻島 和雄[1,a)]

概要：Web ブラウジング中に個人情報の窃取やマイニングマルウェアへの感染等を目的とした悪性コンテンツにアクセスする危険は常に存在する．
これらの対策として，我々はこれまでに URL と遷移情報を利用し，悪性コンテンツを含むページの情報を利用した検知手法及び該当ページへと遷移する１ホップ前のページの情報を利用した事前検知手法の提案を行っている．２つの手法のうち，事前検知の手法は悪性コンテンツに接触する前の段階で警告を発することができるが，検知精度において一段劣るものとなっていた．また，特に１ホップ前が検索サイトである場合に検知精度が低下する問題があった．
本研究では，事前検知において昨年利用した特徴に加えて html 以外で読み込まれる画像，動画やスクリプト等ページ内コンテンツの種別と量に着目し，精度の改善を試みた．加えて，検索エンジンからの遷移に対応するため，データセット内の悪性 URL で学習した分類器によって検索結果 URL の判定を行い，その結果を用いて事前検知結果を補強する手法について検討し，実験を行った．
実験の結果，従来手法より高い精度で悪性コンテンツを含むページの事前検知が可能であることを示した．

キーワード：悪性サイト検知，機械学習

Imorovement of Estimate of Web Content Maliciousness Using Browsing Histry Data

Kazuo Makishima[1,a)]

Abstract: During web crawling, there is always a risk of being exposed to Web-threats.
To deal with these problems, we have proposed the method which use URL and page transition type to detect and pre-detect malicious contents in previous paper. However, there was problem that pre-detection method has lower accuracy than on-spot detection particularly at the case when transition from search site. In this paper, we tried to improve pre-detection accuracy by using in-page contents information. In addition, in order to respond to the transition from the search page, we construct other classifier which learned by malicious URLs in dataset and judge search result URL to reinforce pre-detection.
Experiments shows that using in-page contents and search result URL, we can estimate malicious contents at higher accuracy than the conventional method.

Keywords: Malicious website detection, Machine learning

1. 研究の背景

Web ブラウジング中に悪意のあるコンテンツに遭遇する脅威は，その時々の社会情勢にも応じ様々な事例が報告されている．

大規模なスポーツ大会に乗じた偽配信サイト [1] や感染症に対する調査・情報提供を名目にユーザの個人情報の窃取を試みる例 [2] など，利用者の期待や不安を利用し不正な利益を上げようとする例は多い．

[1] 株式会社セキュアブレイン
SecureBrain Corporation
[a)] kazuo_makishima@securebrain.co.jp

こういった脅威に対応する方法としては，大別して二つのアプローチが考えられる．一つは悪性のコンテンツそのものに着目し，同じ内容のコンテンツや共通した特徴を持つコンテンツを検知しブロックする手法であり，もうひとつは悪性コンテンツの IP アドレスやドメインに着目し，悪性コンテンツへの接続を遮断する手法である．

我々は昨年の論文 [3] において後者のアプローチから悪性コンテンツを検知するため，ページ URL と該当ページへの遷移種別を特徴量とした分類器を構築した．また，悪性コンテンツを含むページ自体の特徴を用いて検知を行うだけでなく，悪性コンテンツを含むサイトに接触する前の段階で次に遷移するページに悪性コンテンツが含まれるか否かを判定する事前検知を試み，検知精度の比較を行った．

昨年の手法では事前検知の条件においても一定の精度を得られたがその精度は対象ページ自体の情報を用いた場合に比べると低く，特に検索エンジンの検索結果ページから 1 ホップで悪性コンテンツを含むページに遷移している場合においては事前検知に使える特徴が良性の場合と悪性の場合で似通っており，精度低下の要因となっていた．

本論文においては事前検知について，昨年の手法に加えページ内コンテンツの情報を特徴量として加えることによる検知精度の向上を試みた．

また，事前検知の精度を低下させる要因となっていた検索結果ページからの遷移パターンについて，ページ URL から取得した検索クエリをもとに再検索を行い，得られた検索結果から抽出した検索結果 URL を用いて該当の状況に特化した分類器を構築した．

更に，検索結果ページから遷移した場合の検知判定を特化した分類器の結果に上書きした場合の精度を検証し，総合的な性能を検証した．

本論文の構成を以下に示す．第 2 章で関連する研究について述べる．第 3 章で利用したデータセットについて収集方法を説明し，収集された内容の分析を行う．第 4 章では昨年の手法をベースラインとしてコンテンツ情報を追加した悪性コンテンツ実験を行い，精度を比較する．第 5 章で検索エンジンから遷移したパターンに特化した分類器の構築を行い，その結果を 4 章のものと統合して評価する．第 6 章で全体の統括と今後の展望を述べる．

2. 先行研究

URL をベースとした悪性サイト検知としては，孫 [4] らによる Baysian Sets を利用し既知の悪性 URL と類似した特徴を持つ URL を探索する研究や，山西 [5] らによるリダイレクトチェインに含まれる URL 群を入力とした研究がある．

検索サイトに関連した悪性サイト検知としては，源平 [6] らによる統計的に平均より危険なサイトへ遷移する率の高い危険検索単語を調査した研究がある．

本研究の先行研究に対する独自性としては、悪性コンテンツを含むサイトそのものではなくその 1 ホップ前の情報を用いた事前検知に主眼を置いていること，検索単語自体ではなく検索を行った結果得られる検索結果 URL を検知基準としていることが挙げられる．

3. データセット

悪性コンテンツを含むサイトにアクセスした際のブラウジング履歴情報を取得するため，WarpDrive 実証実験のデータを利用した．

WarpDrive（Web-based Attack Response with Practical and Deployable Research InitiatiVE/Web 媒介型攻撃対策技術の実用化に向けた研究開発）は，Web 媒介型攻撃の実態把握と対策技術の向上のための研究開発プロジェクト [7] である．

実証実験は 2018 年 6 月より開始しており，Chrome 拡張機能の形で広く一般ユーザに向けタチコマ・セキュリティ・エージェントの配布を行っている．これはブラウザに常駐し危険なサイトへの接続を防ぐと共にユーザの同意を得てブラウジング履歴情報の収集を行うものである．

2020 年 7 月の時点で累計の登録ユーザ数は 1 万を超え，日毎 700 程度のユニークユーザ ID から 1000 万規模のブラウジング履歴情報を収集している．

3.1 データ収集基準

本研究においては収集された WarpDrive 実証実験データより 2020 年 2 月から 2020 年 5 月まで 4 ヶ月分のデータを利用した．

悪性データとしては対象範囲内の実証実験データのうち悪性コンテンツを含むサイトへの訪問履歴を有するものについて，悪性コンテンツへの接触から遡って 20 分間のブラウジング履歴情報を抽出した．

悪性コンテンツを含むサイトの定義としては，ページに含まれるコンテンツのうち一つでも Google Safe Browsing[8] の検知，もしくは WarpDrive の実験基盤で設定したブラックリストで悪性と判定されたものとした．

対照となる良性データとしては，母集団の差異によるデータの偏りを防止するために以下の条件を満たすものからランダムにページへの訪問情報を抽出し，対象となったページから遡って 20 分間のブラウジング履歴情報を抽出した．

良性データの抽出条件

- 対象期間である 4 ヶ月のうちに一度でも悪性コンテンツを含むサイトにアクセスしたユーザである事
- 同日に悪性コンテンツを含むサイトにアクセスしていない日のブラウジング履歴である事

以上の基準により悪性 2860 件，良性 5773 件のブラウジング履歴情報を抽出した．

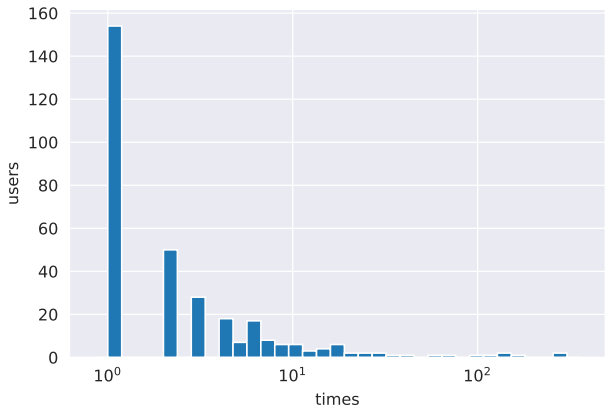

図 1　ユーザ毎アクセス回数

3.2　悪性コンテンツを含むサイトへのアクセスにおける傾向分析

ユーザ毎のアクセス傾向

収集されたデータセット内のユニークユーザ数は 325 であった．同期間中において WarpDrive 実証実験が収集した全データのユニークユーザ数は 2101 であるため，全ユーザのうち約 15%が期間内に少なくとも一度悪性コンテンツに接触している．

悪性コンテンツに接触した経験のあるユーザの半数弱にあたる 154 ユーザは期間中 1 回しか悪性コンテンツに接触していないのに対し，少数ながら反復的に悪性コンテンツに接触しているユーザが存在している．この傾向は昨年の調査と同様であった．

ユーザ毎に，期間中に悪性コンテンツを含むサイトに何回アクセスしたかヒストグラムとして集計した結果を**図 1** に示す．縦軸がカラム毎のユーザ数，横軸が対象ユーザが期間中に悪性コンテンツを含むサイトにアクセスした回数を示す．

頻出ドメイン

収集された悪性コンテンツへの接触について，ドメイン単位で出現頻度の上位を取ると**表 1** のようになる．なお，昨年順位の列においては昨年の同時期（2019 年 2 月から 5 月）に収集されたデータにおける出現頻度順での順位を示す．

傾向として，動画や画像などユーザを誘引するコンテンツを有するサイトが多い．＃動画サイト A，＃動画サイト B については昨年の収集データにおいても上位に出現しており，長期にわたって悪性コンテンツへの接触窓口となっているサイトが存在することが確認できた．

昨年のデータでは見られなかった種類のサイトとして，第 6 位に入っている＃短縮 URL 提供・中継サイトがある．これは短縮 URL を提供するとともに一度このサイトを中継させ，そこで広告等を表示することによって利益を得るサイトのようであった．

表 1　頻出ドメイン名上位 10

悪性サイト	回数	ユーザ数	昨年順位
＃動画サイト A	1291	32	1
＃動画サイト B	248	12	3
＃画像共有サイト	144	1	-
＃動画サイト C	101	1	-
＃動画サイト D	53	6	6
＃短縮 URL 提供・中継サイト	50	17	-
＃動画サイト E	45	5	10+
＃画像サイト	30	4	-
＃ニュースサイト	28	21	-
＃ニュースサイト	24	21	-

アクセス経路

＃短縮 URL 提供・中継サイトについては性質上そのサイト自身が特定の内容を持っているわけではないため，前ホップにあたる遷移元サイトがどのような種類のものであるか調査を行った．

対象サイトへのアクセスを含むブラウジング履歴情報を調査したところ，20 分間の履歴取得範囲内では遷移元サイトを取得できていないケースも多かったが，URL 中に manga といった単語や拡張子.rar が含まれる例があり，漫画作品の電子データをダウンロードさせる際にこういった中継サイトを用いていると推測される．

また，これとは別に著名な・影響力の大きいサイトを経由して悪性コンテンツを含むサイトに遷移している事例が存在するかについて調査を行った．追跡対象として Alexa Top Sites[9] から頻繁にアクセスされる日本のサイト上位 50 件を対象に全ての悪性コンテンツを含むブラウジング履歴から対象のサイトが履歴に含まれるものを抽出したところ，google や yahoo といった検索サイトから遷移しているもののほかに youtube を遷移履歴に含むものが幾つか見られた．

内容を見るとどれも動画説明文にあるリンクを辿った結果悪性コンテンツを含むサイトに誘導されているもので，動画の内容としてはゲームのエミュレータを紹介しているものと Android 端末のブートローダーアンロックの方法を解説しているものであった．

リンクはそれぞれ解説しているツールのダウンロード先として提示されているもので，こういった動画を介して悪性コンテンツに接触するパターンが存在することが判明した．

4.　コンテンツ情報を利用した事前検知実験

データセットの内容にみられるように，悪性コンテンツを含むサイトとして頻出するものには動画系サイトが多い．こういったサイトにはページ構成等に特有の傾向があると考えられるため，URL 以外にページの特徴を捉えることができるような要素を利用することによって検知精度

表 2 content_coarse 分類

分類種別	説明
image	イメージとしてレンダリングされるされるよう読み込まれたリソース
stylesheet	CSS スタイルシート
script	<script>要素によって実行されるコード
font	Web フォント
media	<video><audio>で読み込まれるリソース
xmlhtprequest	xmlhtprequest によって送信されたリクエスト
ping	ping 属性で指定された URL に送信されたリクエスト
object	<object><embed>要素によってロードされるリソース
other	その他

の上昇を図ることができると考えられる．

そこで，悪性コンテンツの事前検知について，昨年の手法に加えページ内のコンテンツ情報を用いた特徴量設計を行い，機械学習の手法を用いて分類器の生成を行った．

前ホップの履歴情報からの特徴抽出を行うため，検知実験の対象データは収集された全データの中から 1 ホップ前の履歴情報が取得できたものに限られる．

履歴の存在するケースを抽出した結果，検知実験の対象となるデータは全悪性データ 2860 件中 1500 件，全良性データ 5773 件中 2196 件，計 3696 件となった．

4.1 実験条件

本実験では全ての条件で対象のページの 1 ホップ前のページに含まれる情報のみを用いて次のページに悪性コンテンツが含まれるか否かを判定する事前検知を行う．

実験条件としては，従来手法に加えページ内コンテンツの分類基準の異なる 2 条件を加え下記の 3 つの条件で実験を行った．

- ページ URL と遷移タイプを利用するもの（従来手法）
- 従来手法に加えコンテンツ情報を属性による粗い分類で用いるもの
- 従来手法に加えコンテンツ情報を形式による細かい分類で用いるもの

記述の簡略化のため以降上記実験条件をそれぞれ baseline，content_coarse，content_fine と表記する．

4.2 利用した特徴量

ページ URL と遷移タイプに関する特徴量の抽出については，昨年の論文に記載したものと同様の手法で行った．URL 特徴量は文字列の長さや含まれる数字の数，特定の区切り文字で分割した際のトークン数など 12 特徴量，遷移タイプに関する特徴量はページ内リンクによる遷移，ページ内のフォームを入力することによる遷移などデータセット内に含まれる全ての遷移タイプについて OneHot ベクトル化した 20 特徴量である．

ページ内コンテンツについては，メインフレームに属すコンテンツ，サブフレーム群に含まれるコンテンツそれぞれについて，内容を問わずにその総数と種類別の出現数を特徴量とした。

コンテンツの種類を分類する基準として，content_coarse ではコンテンツがページ上でどのような役割を果たすかを基に**表 2** に示すよう 9 種類に分類した．

content_fine においては HTTP レスポンスヘッダ中の content-type の値から対象がどういった形式のデータであるかによって分類した．結果，今回のデータセット内で 97 種の形式のものが得られた．得られた例としては audio/mpeg，video/mp4，image/png 等がある．

以上により，条件 baseline では URL による特徴量 12 と遷移情報による特徴量 20 の計 32 次元，条件 content_coarse では baseline に加えて 9 種類の分類別コンテンツ数とコンテンツの総数をそれぞれメインフレーム，サブフレーム群で計数した 20 次元を加え 52 次元，条件 content_fine では同様に 228 次元の特徴量ベクトルが得られた．

4.3 分類器

分類器としてはランダムフォレスト [10] を用いた．これは決定木のアンサンブル学習によって分類を行うアルゴリズムであり，分類にあたって重視された特徴量についての見通しが良く結果の分析が容易なため採用した．

本実験では実装として Python 上で動作する scikit-learn[11] 内の RandomForestClassifier を利用した

4.4 評価基準

分類器の性能を評価する基準としては以下の 4 つを用いる．

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - measure = \frac{2Recall * Precision}{Recall + Precision} \tag{4}$$

数式中の略称について，TP（True Positive）は悪性データを悪性と正しく判定したもの，TN（True Negative）は

表 3 RandomForest パラメタ

パラメタ名	説明	走査範囲
n_estimators	生成する決定木の数	10, 20, ... , 100
criterion	分割純度の算出方法	gini, entropy
max_depth	決定木の深さの限界値	None(Inf), 1, 2, 4, 6, 8, 10
min_samples_split	最小の分割可能ノードサイズ	2, 4, 10, 12, 16

表 4 コンテンツ情報を利用した検知実験 条件毎の最適パラメタ

実験条件	n_estimators	criterion	max_depth	min_samples_split
baseline	90	entropy	None	4
content_coarse	70	entropy	None	4
content_fine	90	entropy	None	2

良性データを良性と正しく判断したものの個数を示す.

同様に，FP（False Positive）は良性データを悪性と誤って判定したもの，FN（False Negative）は悪性データを良性と誤って判定したものの個数を示す.

評価指標のうち，Accuracy は全てのテスト対象データのうち正答したものの割合を示す.

式から見て取れる通り，Precision の高さは FP の少なさ，つまりは誤検知の少ないことを示す．同様に Recall の高さは FN の少なさ，検知漏れの少ないことを示す指標となる.

F-measure は Precision，Recall の調和平均であり，分類器の総合的な性能を示す.

4.5 パラメタの設定

各特徴量セットについて実験条件毎にグリッドサーチにより適したパラメタを設定した．サーチを行った Random-Forest のパラメタとその範囲を**表 3** に，実験条件毎に得られた最良のパラメタセットを**表 4** に示す.

4.6 交差検定

分類器の性能を検証するためデータセットを 5 分割して交差検定を行った．分割内容による偏りを防ぐため交差検定全体を 5 回反復して行い，その平均を結果とした.

4.7 実験結果

実験の結果得られた評価値を**表 5** に示す．なお，search_reinforced の項は 5 章で構築する分類器の結果を用いて補正した場合の値であり，後の章で論じる.

baseline における正答率約 88%に対し，コンテンツ情報を用いた二手法はどちらも正答率 90%を超えており，検知精度の向上がみられる．ページ内コンテンツの情報を使った二手法の間では形式を基にページコンテンツの分類を細かく行った場合のほうが良い結果となった.

評価指標のうち特に Precision が大きく改善しており，新手法は False Positive の抑制において大きな効果を持っていることが見て取れる.

表 5 コンテンツ情報を利用した検知実験結果

実験条件	Accuracy	Precision	Recall	F-Measure
baseline	0.882	0.920	0.777	0.843
content_coarse	0.901	0.970	0.802	0.878
content_fine	0.915	0.980	0.807	0.885
search_reinforced	0.920	0.976	0.822	0.893

4.8 考察

通常のブラウジング中に事前検知の結果を利用して警告を発するといった活用法を想定する場合，一般に訪問するサイトには良性のものが圧倒的に多いことを考えるとユーザの利便性の上で False Positive の抑制は重要であり，今回の手法で Presicion が改善されていることは実用上において有益である.

検知に失敗しているケースについて，baseline と最も高精度であった content_fine で比較すると失敗ケースの数はそれぞれ 435 件と 314 件，双方の条件で失敗している共通部分は 254 件であった.

content_fine でのみ検知に成功しているケースは 181 件であった．対して，baseline で検知に成功していたものが content_fine で失敗しているものも 60 件あった.

content_fine でのみ失敗しているものの内容を見ると，ニュース系のサイトで新たに検知に失敗している場合が多かった．こういったサイトで悪性コンテンツが含まれる場合はページ内の広告が悪性判定されているケースが多く，表示される広告にある程度ランダム性があるため検知が難しいものとなっている可能性が考えられる.

5. 検索結果 URL を用いた事前検知実験

ブラウジング履歴情報のうち，検索エンジンによる検索結果ページから直接遷移しているケースにおいては次ホップのページが悪性コンテンツを含む場合においてもそうでない場合においても似たような特徴量となりやすく，検知精度を下げる要因となっていた.

このようなケースにおいても利用でき，悪性コンテンツ検知の助けとなり得る情報としては URL 中のパラメタから得られる検索クエリがあるが，同一の検索クエリでも時

期等によって検索結果が異なる場合があることや，収集データ内の検索エンジンからの遷移パターンがあまり多くなく検索クエリのみだと十分な教師データを得られないといった問題がある．

そこで，今回の実験では検索クエリを元に再度検索を実行し，得られた検索結果 URL を特徴量として検索結果ページからの遷移について事前検知実験を行った．

5.1 対象となるケースの抽出

本実験においては検索エンジンから遷移しているパターンのうち google 検索エンジンから遷移しているものを対象とする．

データセット中 1 ホップ前が google による検索結果であったパターンは悪性 67 件，良性 108 件の合計 175 件存在した．

5.2 検索結果の再取得

175 件の URL からパラメタ q=の部分を抽出し，検索クエリを得る．得られた検索クエリを元に改めて検索を行い，結果の 1 ページ目に含まれるリンク URL を取得した．

今回の実験では 2020 年 2 月から 5 月のブラウジング履歴情報について 2020 年 7 月に再検索を行った．

検索の結果，検索クエリ 1 件当たり平均 10.14 件，計 1774 件の検索結果 URL が得られた．

理想的にはそれぞれの検索クエリから得られた検索結果 URL の中に取得元のデータにおいて次に遷移しているページの URL が含まれるはずであるが，履歴の取得にある程度の時間的なずれが存在することや google がパーソナライズドされた検索を行っていることにより，取得元のデータにおいて遷移先ページ URL が再検索時には現れない場合も多くみられた．

今回取得した結果においては，175 件中 80 件の検索結果において取得元のデータにおける遷移先ページ URL が検索結果に現れていた．残りの 95 件については元のデータにおける遷移先とは異なる URL しか得られなかったが，同一の検索クエリから得られた結果にはある程度の類似性があると考えられるため，得られた結果をそのまま検知実験に用いた．

5.3 教師データ

検索結果 URL は単一の検索クエリから複数の URL が得られるため，悪性側のデータから得られたものであってもその全てが悪性コンテンツを含むページの URL というわけではない．このことは良性側についても同じことが言える．

それゆえ，検索結果 URL をそのまま分類器の学習データとして利用することは不適当である．

そこで，教師データとしてはデータセット全体（8633 件）における悪性コンテンツを含むページの URL/対象となる悪性コンテンツを含まないページの URL をそれぞれ悪性サンプル，良性サンプルとして利用する．

5.4 特徴量抽出

4 章の実験においては URL からの特徴量抽出において文字数をベースとした抽象的な手法を用いたが，別角度からの特徴を組み合わせることによる精度の向上を図るため本実験においては URL を特定の区切り文字により分割したうえで，分割されたトークンによる Bag-of-Words をとり，特徴ベクトルとする..

区切り文字としては [“&”, “%”, “/”, “?”, “=”, “-”, “_”, “.”] の 8 種を採用する．

Bag-of-Words の生成にあたっては，全 URL 中 80%以上で出現しているトークンについては汎用的なものであり特徴に寄与しないとみなして除去を行った．

結果，次元数 22030 の特徴ベクトルが得られた．

5.5 分類器

分類器としては第 4 章の実験と同じくランダムフォレストを用いる．

5.6 評価基準

ランダムフォレストに教師データから得られた特徴量ベクトルを入力し，分類器の訓練を行う．

評価は 175 件の google 検索エンジンから遷移したパターンのデータに対しそれぞれ得られた検索結果 URL について行われるが，一件のデータから複数の検索結果 URL が得られる為，単純に URL ごとの良性/悪性判定精度を検知精度とすることはできない．

よって，今回は得られた検索結果 URL のうち一件でも悪性と判定された場合対象のデータは悪性と判定されたものとみなすこととした．

5.7 実験結果

実験の結果得られた評価値を**表 6** に示す．search_result の項が今回の実験による事前検知の結果である．

baseline と content_fine の項は 4 章で行った実験により得られた分類器で対象となった 175 件のデータについてそれぞれ判定を行った場合の精度である．

baseline においては検索結果ページからの遷移パターンについて無作為よりはやや良い，といった程度の検知精度であった．content_fine では Precision が 1 を記録しており False Positive は出ていないものの，検索サイトからの遷移全般を良性に判定する傾向が強く出ており Recall が低いため，F-measure は低い水準にとどまっている．

検索結果 URL を使う手法では Precision と Recall がバランスよく高い値となり，全体の正答率も増している．

表 6　検索結果の再取得による検知実験結果 n=175

実験条件	Accuracy	Precision	Recall	F-Measure
search_result	0.897	0.915	0.806	0.857
baseline	0.680	0.667	0.328	0.440
content_fine	0.800	1.000	0.478	0.646

5.8　実験結果を利用した悪性コンテンツ事前検知の補強

対象となった 175 件のデータについて，search_result の結果によって content_fine の結果を上書きした時の事前検知実験全体の 3696 件に対する検知精度を表 5 に示す．

データセット全体に対する検索結果からの遷移パターンによるケースの数が小さいこともあり変化は少量にとどまっているものの Accuracy，Recall の数値が上昇している．

5.9　考察

再検索によって得られた URL はそのすべてが悪性コンテンツを含むサイトへのリンクというわけではなく，特に今回は半数弱がもともと遷移していた URL を収集できていない条件の下での実験であった．

しかし、検知精度としては正答率で 9 割程度と低くはない数値になっている．

検索結果 URL のうち取得元のデータにある URL が得られなかったものについてどのような検索クエリによるものだったかを調査してみたところ，特定時期に行われたイベント等についての検索だったが再検索の時期が離れていたために取得できなかったものも見られたが，多くは作品名等について調べた際その作品に関するサイトが多数存在するために上位に元の URL が含まれなかったというパターンのようであった．

特に悪性側のケースにおいて映像作品名で検索しているパターンがあり，ほかの動画サイトでも同じ作品がアップロードされた結果元 URL が再現されなくなるという場合が多く見られたが，こういった場合 URL にパーセントエンコードされた作品名が含まれるケースが多く，元 URL とある程度に通った特徴を持ったものが検索結果 URL として取得できているため精度を低下させる要因とはならなかったものと考えられる．

content_fine の条件だけでも baseline に比べて検索サイトからの遷移について精度が上がっている．先程の URL にパーセントエンコードされた文字列が含まれるという特徴は (悪性コンテンツを含むことが多い) 動画系サイトと URL に検索クエリ情報を含む検索結果ページで共通するものであり，baseline の手法では混同されがちであった．ページ内コンテンツの情報を用いることでそれらを区別して検知ができるようになり，精度が上がったものと考えられる．

6.　まとめと今後の展望

コンテンツ情報を用いた事前検知実験では従来手法に比べ高い精度で悪性コンテンツの含まれるページへの事前検知を行うことができ，特に False Positive を大幅に減らすことができた．

さらに、検索結果 URL を利用した分類器を組み合わせることによって全体からの割合という点では少量であるが以前の手法で欠点となっていたパターンについて検知精度を改善することに成功した．

以降の展望としては，検索結果 URL の取得について google 以外の検索エンジンを対象に加えることで適用範囲を広げることや，検索結果の再取得までのラグを減らして精度の向上を図ることが考えられる．

また，WarpDrive 実証実験環境を利用して今回得られた分類器を実地に適用し，悪性コンテンツを含むサイトへの遷移が考えられる場合事前にユーザに対して警告を発する仕組みの構築などを試みたい．

謝辞　本研究は，国立研究開発法人情報通信研究機構の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」の成果の一部です．ご協力いただいた皆様に，深く感謝します．

参考文献

[1] 【注意喚起】ラグビーワールドカップ人気に便乗したフィッシング詐欺に注意，TRENDMICRO https://is702.jp/news/3568/partner/97_t/
[2] グローバルセキュリティ動向四半期レポート pp3-10，NTTDATA https://www.nttdata.com/jp/ja/-/media/nttdatajapan/files/news/information/2020/062600/062600-01.pdf
[3] 巻島 和雄, 三須 剛史：ブラウジング履歴情報に基づく悪性サイトの事前検知, Computer Security Symposium 2019
[4] 孫 博, et.al：既知の悪性 URL 群と類似した特徴を持つ URL の検索, Computer Security Symposium 2014
[5] 山西 宏平, et.al：畳み込みニューラルネットワークを用いた URL 系列に基づくドライブバイダウンロード攻撃検知, Computer Security Symposium 2016
[6] 源平 祐太, et al：悪性 Web サイトに到達しやすい危険検索単語の検知, Computer Security Symposium 2019
[7] WarpDrive https://warpdrive-project.jp/index.html
[8] Google Safe Browsing https://safebrowsing.google.com/
[9] Alexa The top 500 sites on the web https://www.alexa.com/topsites/countries/JP
[10] Breiman, Leo (2001). “Random Forests”. Machine Learning 45 (1): pp5?32
[11] scikit-learn https://scikit-learn.org/stable/