Vol.2018-DPS-174 No.5
Vol.2018-CSEC-80 No.5
2018/3/5

機械学習を用いた脅威インテリジェンス抽出手法

林翔太[†1]　辻秀典[†2]　橋本正樹[†3]

概要：近年，サイバー攻撃の手法は日に日に高度化しており，防御側で対策を採っていても，攻撃を完全に防ぐことは困難となっている．この状況を打開するためには，サイバー攻撃を予測し，事前に適切な対応を行うことが必要であり，これを可能とするインテリジェンスの活用が重要となる．一般に，攻撃者の多くは，ダークウェブや特殊なコミュニティにおいて攻撃に利用可能な情報やツールを共有しており，サイバー空間には，これらを含む多くのインテリジェンスが眠っているものと推測できる．そこで，本研究では，ダークウェブ上に存在する様々なフォーラムに着目し，フォーラム内の投稿に機械学習を適用することで，重要情報を含むフォーラムを抽出するとともに，フォーラムの特性を明らかにすることを目指す．これにより，サイバー空間上の脅威情報を適時適切に把握し，事前に最適な防御策を講じることが可能となることを期待するものである．

キーワード：インテリジェンス，ダークウェブ，doc2vec，機械学習

Exploring Darkweb for Cyber Threat Intelligence using Machine Learning

SHOTA HAYASHI[†1]　HIDENORI TSUJI[†2]
MASAKI HASHIMOTO[†3]

Abstract: In recent years, cyber attack techniques are increasingly sophisticated, and blocking the attack is more and more difficult, even if a kind of counter measure or another is taken. In order for a successful handling of this situation, it is crucial to have a prediction of cyber attacks, appropriate precautions, and effective utilization of cyber intelligence that enables these actions. Malicious hackers share various kinds of information through particular communities such as dark web, indicating that a great deal of intelligence exists in cyberspace. This paper focuses on forums on darkweb and proposes an approach to extract forums which include important information or intelligence from huge amounts of forums and identify traits of each forum using methodologies such as machine learning, natural language processing and so on. This approach will allow us to grasp the emerging threats in cyberspace and take appropriate measures against malicious activities.

Keywords: Intelligence, Darkweb, doc2vec, Machine Learning

1. はじめに

近年, サイバー攻撃の手法は, 日に日に高度化している．かつて，サイバー攻撃は，いたずらを目的とした個人的なものが多数を占めていたが，現在では，金銭詐取を目的とした組織化されたものが増加している．また，かつて，サイバー攻撃の対象は無差別であったが，近年は，特定の対象に対して, 特定の目的を持って執拗に攻撃を仕掛ける「標的型攻撃」が主流となっている．さらに，セキュリティ調査機関 AV-TEST の調査によると[1]，2016 年には毎日およそ 35 万の新種のマルウェアが製造されていたという．

こうした最近の情勢により，サイバー攻撃に対して防御側で対策を採っていたとしても，全ての攻撃を完全に防ぐことは困難になっている．防御側は，いわば“防戦一方”の状況に陥っていると言えるのである．

この状況を打開するためには，サイバー攻撃を予測し，事前に適切な対応を行うことが必要であり，これを可能とするインテリジェンスの活用が重要となる．一般に，攻撃者の多くは，ダークウェブや特殊なコミュニティにおいて攻撃に利用可能な情報やツールを共有しており，サイバー空間には，これらを含む多くのインテリジェンスが眠っているものと推測できる．インテリジェンスを活用することにより，事前に攻撃を察知し，“積極的防御”を展開するのである．

本研究の目的は，自然言語処理の一手法である doc2vec 及び機械学習を駆使することで，ダークウェブの投稿から「先行的な対策を講じるにあたり有用な情報を含む投稿」（以下，重要投稿という）を抽出し，この結果をもとに，ダークウェブに存在する様々なフォーラムの特性を明らかにすることである．

ダークウェブには，多くのフォーラムが存在している．悪意あるハッカーらは，こうしたフォーラムにおいて，マ

†1 情報セキュリティ大学院大学
Institute of Information Security.
†2 同上
†3 同上

ルウェアの売買やハッキング技術に関する情報を投稿し共有を図っている．これらの投稿に doc2vec 及び機械学習を適用することで，重要投稿を抽出し，さらに，抽出した重要投稿との関連から，これらの投稿のプラットフォームとして機能するフォーラムの特性を詳らかにすることを目指す．すなわち，各フォーラムにどのような重要投稿が含まれているかが明らかになれば，各フォーラムの特性を明らかにすることが可能となるであろうことを想定している．

2. 背景

本章では，インテリジェンス，ダークウェブ，doc2vec 及び機械学習について述べ，本研究の背景を示す．

2.1 インテリジェンス

インテリジェンスと似て非なるものに，インフォメーション(情報, information)がある．インフォメーションとは，それがいかに発見されるかにかかわらず，我々が知りうる全てを指す．他方，インテリジェンスとは，受領者の需要として明確にされたもの，またはそのような需要と理解されたものに合致する情報であり，当該需要に応えるため，収集され，処理され，絞り込まれた情報である[2]．すなわち, インテリジェンスは, 情報から生成されるものであり，その生成のプロセスは，収集，処理等からなる．そして，インテリジェンスは，受領者の要求に沿うものでなければならない．

現在，サイバー攻撃への積極的防御としてインテリジェンスを活用する動きが, 企業を中心に広がりを見せており，その中には目覚しい成果を上げているものもある[3][4][5]．他方，ダークウェブには，インテリジェンスとして活用できる投稿だけが存在しているわけではない．ドラッグやポルノ，雑談，デマ情報など，膨大な数の投稿が飛び交っており，インテリジェンスとなり得る情報は，これらの雑多な情報の中に埋もれている．

インテリジェンスには適時性が求められることを合わせて勘案すると，玉石混淆の情報が溢れるダークウェブから効率的に重要投稿を抽出することができれば，セキュリティ分析者が生成するインテリジェンスは，適時性を兼ね備えた強力なものとなる．すなわち，提案手法によって，サイバー空間上の脅威情報を適時適切に把握し，事前に最適な防御策を講じることが可能となることを期待するものである．

2.2 ダークウェブ

インターネットには階層が存在する．通常，我々が目にするウェブコンテンツは, サーフェスウェブ（Surface Web）という階層に属することが多い．例えば，Google や Yahoo などの検索エンジンで発見可能なコンテンツが挙げられる．他方，インターネットの“深部”には，通常の検索方法ではアクセスできないコンテンツが無数に存在している．これらの階層は，ディープウェブ（Deep Web）及びダークウェブ（Dark Web）と呼ばれる．

ディープウェブとは，様々な理由で通常の検索エンジンではインデックスされないウェブコンテンツを指す．例として，登録制のウェブサイトが挙げられる．また，ダークウェブとは，ディープウェブの一部で，意図的に隠されたウェブコンテンツを指す[6]．

ダークウェブは，Tor 等の特殊なソフトウェアを使用しなければアクセスできず，極めて匿名性が高い領域となっている．ダークウェブには，無数のフォーラムやマーケットが存在しており，これらが，あらゆる犯罪の温床となっていることが指摘されている[7]．犯罪者等は，薬物，銃，ポルノ，盗品といった通常の市場では扱うことができない商品をダークウェブで取引しているのである．

特に近年，ダークウェブの取引に変化が見られている．取引の中心が，薬物等からマルウェア，機密情報，個人情報，医療記録，ハッキングのチュートリアル，クレジットカード番号，金融機関の口座といった商品にシフトしているのである[8]．こうした商品は，サイバー空間上の脅威の副産物とも言える．したがって，ダークウェブからの情報収集・分析の精度を高めることは，サイバー攻撃への“積極的防御”として有効な手立てとなりうるものと考えられる．

2.3 doc2vec

本研究では，ダークウェブの投稿を研究対象とする．投稿はテキストデータであるため，機械学習のインプットとして使用するためには自然言語処理を行わなければならない．提案手法では，doc2vec を使用する．

現在, 文書をベクトル化する手法としては, Bag-of-Words（BoW）が広く使用されている．BoW は，文書中の各単語の出現頻度を特徴量として，テキストをベクトル化する手法である．同手法は，単純に文書中の各単語の出現頻度をカウントすることで特徴量を抽出し，単語が出現する前後関係を考慮しないことから，単語の意味を表現することができないという課題を抱えている．この課題を克服するのが doc2vec である．

doc2vec は，米グーグル社の研究者である Tomas Mikolov が提案した自然言語処理の手法[9]で，ニューラルネットワークを利用して文書をベクトル化する．doc2vec は，単語の前後関係を考慮したモデルであることから，文書中の各単語に意味を持たせた形でベクトル化することが可能となっている．

doc2vec は，単語レベルでの高精度のベクトル化を可能とした word2vec[10]の理論を文書レベルに応用したものである． word2vec には，CBoW（Continuous Bag-of-Words）

及び skip-gram の 2 つのアルゴリズムがあり，ぞれぞれ，doc2vec の PV-DM（Paragraph Vector-Distributed Memory）及び PV-DBoW（Paragraph Vector-Distributed Bag-of-Words）に対応する．

2.4 機械学習

本研究では，doc2vec で抽出した特徴量をインプットとして機械学習を行う．以下では，機械学習の基礎について説明する．

機械学習（Machine Learning）は，人工知能の研究から派生した研究分野で，抽象的なモデルをデータから学習するための方法論である．

機械学習は，一般に学習方法の観点から，教師あり学習（Supervised Learning）及び教師なし学習（Unsupervised Learning）に大別される．教師あり学習は，入力データと望ましい出力のペア群をアルゴリズムに与え，その規則性を学ぶ学習方法である．教師なし学習は，望ましい出力をアルゴリズムに与えず，データの背後に存在する本質的な構造を抽出する学習方法である．

さらに，教師あり学習は，分類及び回帰に大別することができる．分類（classification）の目的は，選択肢の中からクラスラベルを識別することである．スパムメールの判別は，分類の一例である．回帰（regression）の目的は，連続値を予測することである．天候データから農作物の収穫量を予測することは，回帰の一例と言える．また，教師なし学習は，クラスタリングに代表される．クラスタリング（clustering）の目的は，類似しているデータ群をクラスタ毎に分割することである．クラスタリングは，分類の前段階として，探索的な手法として用いられることが多い．

本研究では，教師あり学習の分類を用い，重要投稿とそうでない投稿を峻別する．

3. 関連研究

本章では，関連研究の概要及び課題について述べる．

3.1 関連研究の概要

Eric Nunesら[11]は，ダークウェブ上のマーケットやハッカーらが集うフォーラムから，ユーザーの個人情報や脆弱性，ハッキング技術といったサイバー攻撃に関係すると思しき情報を，機械学習を駆使して抽出する研究を行った．通常，教師あり学習においては，データのラベル付けに大幅な時間と専門的知見が要求される．Nunesらは，半教師あり学習により，この課題を克服することを試みた．本研究は，Eric Nunesらの研究をベースとしている．

Sagar Samtaniら[12]は，データ解析技術である Latent Dirichlet Allocation(LDA)と機械学習を組み合わせた研究を行った．LDAは，文書のトピックを確率的に求める言語モデルである．Samtaniらは，地下のコミュニティに投稿されているソースコード等に着目し，LDAを用いてこれらの機能やトピックを抽出し，サイバー攻撃に関連するソースコードの抽出を試みた．さらに，Samtaniらは，ソースコードに使用されているプログラミング言語に着目し，機械学習を駆使してソースコードとプログラミング言語の相関性を明らかにした．

Ahmed Abbasiら[13]は，機械学習のクラスタリング（教師なし学習）を活用した研究を行った．従来の研究が，影響力を持つハッカーの存在を明らかにすることに主眼が置かれていたのに対し，同研究では，ハッカーの特性や専門性を明らかにした点が注目される．

Victor Benjaminら[14]は，情報検索技術と機械学習のクラスタリング（教師なし学習）を組み合わせた独自のシステムにより，脅威インテリジェンスや脆弱性情報を抽出する手法を提案している．Benjaminらは，フォーラムに加え，“Internet-Relay-Chat(IRC)” にまで研究対象を広げた．IRCを含めることにより，時宜性の高いインテリジェンスを抽出することが可能となる．Benjaminらは，重み付けされたキーワード検索技術の活用を主眼としている．

Mitch Macdonaldら[15]は，Sentiment Analysisを活用した研究を行い，フォーラムの投稿に内在する感情と，同感情の矛先となる重要インフラとの関係を明らかにした．Sentiment Analysisは，与えられたテキストの背景にある感情的な考え方を分析する手法である．Macdonaldらは，品詞タグ付け等の技術を活用し，サイバー攻撃関連の単語及び重要インフラ関連の単語を含む投稿を抽出し，抽出された投稿に対してSentiment Analysisを行った．

3.2 関連研究の課題

従来の研究は，大きく 2 つに分けることができる．第 1 に，フォーラム上の影響力の大きいユーザーの発見やユーザーの特性解明を目的とした，ユーザーを軸に据えた研究である．第 2 に，フォーラムの投稿の中身を分析し，脅威インテリジェンスを直接抽出する，投稿を軸に据えた研究である．これらの研究は，インテリジェンス抽出のために有益なものであるが，課題もある．まず，第 1 の研究に関しては，ユーザーは，頻繁にアカウントを変更することが多い．そのため，ユーザーの流動性という課題に対処しなければならない．第 2 の研究に関しては，あらゆる投稿を無差別に研究対象にした場合，投稿の真実性に疑問符がつく場合がある．この場合，信頼できるソースから得た投稿であることが要求される．

本研究では, 攻撃者が活動を行うメインステージであり，多くの投稿が集まるダークウェブのフォーラムに焦点を当てる．現状，ダークウェブには，無数のフォーラムが存在し，それぞれのフォーラムがいかなる特性を持っているかについては研究が進んでいない．サイバー上の各種脅威に

特化したフォーラムの存在が明らかとなれば，情報収集の効率性及び確度は，格段に向上するものと思われる．フォーラムは，ユーザーアカウントと比較し存続期間が長いため，ユーザーの流動性という課題を克服することが可能となる．また，各種脅威に特化したフォーラムは，信頼できるソースとなり得ることから投稿の真実性という課題にも対処することが可能となる．

4. 提案手法

本章では，本研究の提案手法である「機械学習を用いた脅威インテリジェンス抽出手法」の概要について述べる．

4.1 提案手法概要

本研究では，自然言語処理の一手法である doc2vec 及び機械学習を駆使し，ダークウェブの投稿のうち，重要投稿とそうでないものを自動的に判別する．さらに，この結果を利用し，ダークウェブに存在する様々なフォーラムの特性を明らかにする．具体的には，図 1 の手順で処理を行う．

以下，各手順について説明する．

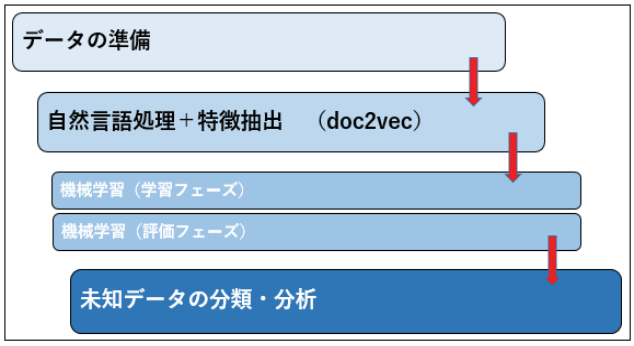

図 1　提案手法の手順

Figure 1 Flow chart of the proposal

4.2 学習データの準備

機械学習の教師あり学習を行うためには，相当量のデータを収集し，さらに収集したデータを正解データと不正解データにラベル付けし教師データを作成する必要がある．

提案手法では，ダークウェブ上のデータ収集に特化したウェブクローラー等を使用してデータを収集する．本稿の実験では，Sixgill という有償のツールを使用した（後述）．投稿データを収集した後，重要投稿とそうでない投稿を手動で峻別し，前者を正解データ，後者を不正解データとして扱う．なお，本稿の実験では，重要投稿として，「マルウェア・オファーに関する投稿」を設定した（後述）．

4.3 自然言語処理・特徴抽出

フォーラムの投稿はテキストデータであるため，機械学習の入力とするために自然言語処理を施し，また，機械学習の前段階として，分類に適した特徴量を抽出する必要がある．提案手法では，自然言語処理及び特徴抽出に doc2vec を使用する．ここで得られる特徴量は，各単語に意味を付与し，文脈を考慮したものとなっている．

文書を適正にベクトル化し，特徴量を得るためには，自然言語処理の前処理が決定的に重要となる．そこで，提案手法では，前処理として，単語のトークン化，クリーニング処理，単語の正規化，Stemming 処理，Stop-Words 処理を行う．単語のトークン化は，単語間の区切りを明確化する処理である．クリーニング処理は，テキスト中の数字や()といった余計な文字を削除することを指す．単語の正規化では，大文字小文字を統一する．Stemming 処理は，派生語などを同一の素性とみなす処理である．Stop-Words 処理により，“I”, “Have”といった，どんな文章にも出現し，タスクを解くのに不要な単語を削除する．

4.4 機械学習

doc2vec により自然言語処理を施し，特徴量を得た後には，同特徴量をもとに機械学習を実施する．

機械学習は，モデルを生成する学習フェーズ及びモデルの性能を評価する評価フェーズからなる．学習フェーズでは，訓練データを用いて doc2vec によって得られた特徴量を学習し，モデルを生成する．評価フェーズでは，評価データによりモデルの性能を評価する．データを訓練データと評価データに分割するのは，未知データに対するモデルの汎化性能を測るためである．

提案手法では，教師データを無駄なく使用するため，k 分割交差検証（k-fold Cross Validation）を実施する．k 分割交差検証とは，訓練データを k 分割し，分割したデータ集合の 1 から k－1 までの集合から学習して，k 番目のデータ集合をテストし，組み合わせを k 通り全て試す方法である．

4.5 未知データの分類・分析

モデルの汎化性能が確保された後は，ダークウェブの未知データをモデルに投入する．ここでは，複数のフォーラムの投稿をモデルに投入し，各フォーラム中の重要投稿を抽出する．その後，各フォーラムの重要投稿の総数や全投稿数に占める割合などを比較することにより，いずれのフォーラムが重要投稿を多く含むかをランキング表示する．表 1 は，ランキングの例である．

これにより，「○○○（インテリジェンスの種類）に関しては，○○○（フォーラム名）を監視する」などといった，セキュリティ分析者にとってのインテリジェンスを抽出することが可能となる．さらに，上位に位置するフォーラムをピンポイントで監視することにより，効率的かつ質の高い情報収集が可能となり，洗練されたインテリジェンスの抽出が可能となる．

表 1　データ分析の例
Table 1　Example of data analysis

順位	フォーラム
1	Exploit.in
2	silkload
3	hackfive
4	hackhound
5	icode
6	Alpha Bay
7	code blue
8	gumtape
9	anonymous
10	malmarket

5.　実験

本研究の提案手法の有効性を実証するため，以下の実験を行った. 提案手法では, 重要投稿の抽出を目指しており，ここでは，これを「マルウェア・オファーに関する投稿」と設定した．実験は，図 1 の手順に従い実施した．

5.1 実験の環境

実験は，表 2 の環境下で実施した．

OS	Windows 7 Professional
CPU	Intel® Core™ 2 Duo CPU
メモリ	3.00 GB
システムの種類	32 ビットオペレーティングシステム
プログラミング言語	python
開発環境	Pycharm

表 2　実験の環境
Table 2　Experimental environment

5.2 データの準備

実験データの収集には Sixgill を使用した．Sixgill は，ダークウェブ上で活動するハッカーの活動情報や, SNS 情報，組織内のヒエラルキー分析を可能とするインテリジェンスプラットフォームである[16]．収集対象は，英語の投稿とした．

Sixgill は，高度なキーワード検索機能や投稿を doc 形式でテキスト化するエクスポート機能を有している．これらの機能を用い, さらに目視で各投稿を確認することにより，「マルウェア・オファーに関連する投稿」(正解データ) 850 件及び「『マルウェア・オファーに関連する投稿』とは無関係の投稿」850 件を用意した．表 3 は，正解データと不正解データの例である．

表 3　ラベル付けの例
Table 3　Examples of Posts label

正解データ
Hello Everyone i am selling one of the crypter that was once sold on HF i bought the source code from the coder but now am busy with another project am doing so i will like to sell it for anyone who is interested in it.
不正解データ
On one hand the judiciary is giving verdicts against the poll body one day to the election. On the other hand, the poll body is continuing to prepare for tomorrows election. there is a case to be heard and determined by the SC involving all presidential candidates, the poll body, its chairman, and 3 petitioners.

5.3 自然言語処理・特徴抽出

まず，適正なベクトル化のために，単語の正規化，単語のトークン化，Stop-Words 処理，Stemming 処理及びクリーニング処理を行った．

次に，doc2vec のモデルを構築し，テキストをベクトル化した．doc2vec は，python のライブラリである gensim に用意されている[17]．前述のとおり，doc2vec には 2 つのモデルがあるが，実験では，デフォルトの PV-DM を使用した．PV-DBoW が語順を考慮しないモデルとなっているのに対し，PV-DM は，語順を考慮したより精度の高いモデルになっているためである．さらに，doc2vec では，学習の中では決定されないハイパーパラメータを設定する必要がある．ここでは，数度の試行を経て，次元数を 200 次元，学習の繰り返し回数を 300 回と設定した．

5.4 機械学習

doc2vec で得た特徴量をもとに機械学習を行った．実験では，SVM（Support Vector Machine）及びニューラルネットワークの一種である MLP（Multi-Layer Perceptron，多層パーセプトロン）を使用した．SVM 及び MLP のアルゴリズムを始め，以下の実験で登場する機械学習の手法は，いずれも python のライブラリである scikit-learn[18]に用意されている

通常，機械学習の入力となるベクトルは，0～1 や-1～1 といったように一定の範囲内に収める必要がある．際限がなければ値が分散してしまい，同じ尺度で学習できないためである．doc2vec により得られたベクトルは，値が分散しているためスケーリングを施した．スケーリングは，ベクトルのとり得る値の範囲を調整する手法である．これによりベクトルは，概ね-1～1 の値に収束した．

さらに，SVM 及び MLP は，いずれも学習の中では決定されないハイパーパラメータを設定する必要があるため，グリッドサーチによりパラメータの最適化を行った．グリ

ッドサーチとは，ハイパーパラメータの探索空間を格子状に区切り，交点となるハイパーパラメータの組み合わせについて網羅的・自動的に試行する手法である．

評価手法としては，層化 k 分割交差検証を取り入れた．層化 k 分割交差検証（Stratified k-fold Cross Validation）は，訓練データと評価データにおける各ラベルの比率が均等になるようにデータを分割する手法である．実験では，正解データを 0，不正解データを 1 と設定し，ディレクトリの上から順に[0, 0, 0,……,1, 1, 1]とラベル付けしているため，通常の k 分割交差検証では，訓練データと評価データにおけるラベルの比率に著しい偏りが生じる可能性がある．層化 k 分割交差検証を使用することにより，上述の問題を解決することが可能となる．実験では，k=10 とした．評価指標は，精度（Accuracy）を採用した．図 2 は，SVM の実行結果である．

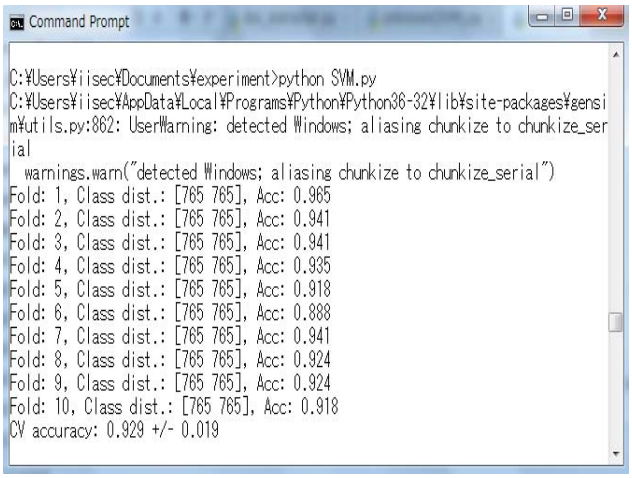

図 2　SVM 実行結果

Figure 2　SVM Classification Performance

10 回の交差検証において，正解データ[0]765 件及び不正解データ[1]765 件の合計 1,530 件で学習を行い，各データの 85 件ずつ合計 170 件で評価を行っている．最終的に，各検証の平均と標準偏差が出力されている．

精度（Accuracy）0.929，標準偏差+/-0.019 という結果を得た．実行時間は，約 10 秒であった．

なお，実験では，SVM が MLP の精度を上回ったため，ここでは SVM の実行結果のみを示した．

5.5 未知データの分類・分析

前項で構築した SVM のモデルを利用して，未知データの分類・分析を試みた．

未知のデータとして，ダークウェブの dreammarket，hackforums 及び kenyatalk という 3 つのフォーラムから 500 件ずつ投稿を収集した． Sixgill は，投稿総数をもとに，「活発に活動している」フォーラムをランク付けしている．これらのフォーラムは，データ収集時（2017.12）に同ランキングの上位に位置していたものである．表 4 は，分類結果をまとめたものである．

表 4　分類結果の比較

Table 4 Results of classification comparison

		ラベル		マルウェア・オファーに関連する投稿の割合
		0（マルウェア・オファーに関連する投稿）	1（無関係の投稿）	
フォーラム	dreammarket	255	245	51%
	hackforums	386	114	77%
	kenyatalk	283	217	56%

分類結果をもとに，仮にフォーラムをランク付けすると表 5 のようになる．表 5 は，全投稿数（実験データ）に占めるマルウェア・オファーに関連する投稿の割合を比較したものである．同表の正確さについて，次節で考察を試みる．

表 5　マルウェア・オファーのプラットフォーム

Table 5　Offering malwares platform rankings

順位	フォーラム
1	hackforums
2	kenyatalk
3	dreammarket

5.6 考察

表 6 は，未知データのほとんどを占める「『マルウェア・オファーに関連する投稿』とは無関係の投稿」である．これらは，全て 0（マルウェア・オファーに関連する投稿）と判定された．実験で使用した未知データは，ダークウェブから無差別に投稿を収集したものである．したがって，「マルウェア・オファーに関連する投稿」は，ごく限られた数しか含まれていないものと思料される．実際，未知データの投稿を目視で確認したところ，「マルウェア・オファーに関連する投稿」は，数えられる程度であった．

しかし，実験で生成したモデルは，いずれのフォーラムも「マルウェア・オファーに関連する投稿」を 50%以上含むと判定した．これは，表 6 から明らかなとおり，多くの誤判定の結果であると考えられる．また，表 6 の投稿を見ると，少なくとも目視で確認する限りにおいては，各投稿に共通性が認められるものではない．したがって，実験において SVM により生成したモデルが，何をもってこれらの投稿を 0（正解データ）と判定したかについては定かで

ない．

以上から，実験で生成したモデルは，未知データの分類に関しては対応できていないと言わざるを得ない．

表 6　未知データ中の不正解データ
Table 6　Negative data within the unknown data

投稿
Are you guys getting the DDOS captcha page if you do manage get onto a market mirror? Thanks,
anyone know if there are any white hats doing services to detect any malware rats etc? Im looking but I can't find although Im sure somethings up. Random black CMD boxes pop up and now getting emails from services stating accounts logged in from unkown locations.
Holy Father I thank you this evening. Thank you for being with us the whole day and bringing us back to our houses. Forgive us where we have wronged knowingly or and unknowingly. This night Father God, I pray that you may cover us by your blood. Surround us with your Angels Jehovah. Let us not soak our pillows with tears this night Lord but heal our broken hearts and fulfill your promises to us. Answer every secret prayer from everyone here and let it be a testimony in Jesus Mighty Name. Mark register

他方で，同モデルは，層化 k 分割交差検証において平均精度が 90%以上という高い汎化性能を示していた．この矛盾について，主に 2 つの原因が考えられる．

第 1 に，データ数が足りていないことである．実験において使用した教師データの数は，1,700 件と決して多いものではない．このため，機械学習のモデル構築の際に，「マルウェア・オファーに関連する投稿」に内在する規則性を捕捉できなかった可能性がある．したがって，モデルは，いわば「少ない訓練データの中に内在する規則性」を学習し，これについての評価を実施したため，層化 k 分割交差検証では高い汎化性能を示しながら，未知データでは，「マルウェア・オファーに関連する投稿」と無関係の投稿の分類というタスクをうまく判定できなかったものと考えられる．

第 2 に，「マルウェア・オファーに関連する投稿」と無関係の投稿の“中間”に位置する投稿に対し，モデルが対応できなかった可能性がある．“中間”に位置する投稿とは，例えば，「マルウェアに言及するもののオファーでない投稿」，「クレジットカード情報の売買に関する投稿」などである．これらは，実験における「データの準備」において，始めに正解データを収集するためにキーワード検索でヒットした 2,000 件のうち，「関係しそうであるが，無関係である」として除外したものの多くが該当するであろう．これらの投稿を不正解データとして教師データに入れ，モデルの構築に際し，アルゴリズムが“中間”に位置する投稿を不正解と判定できるように学習させる必要があったものと考えられる．

以上が実験に関する考察である．実験で構築したモデルは，未知データには対応できなかったものの，未知データの分類に向けた方法論を如実に示している．実験の考察を踏まえ，モデルを改善することで，未知データの分類，より具体的には，ダークウェブから重要投稿を抽出し，フォーラムの特性を明らかにすることが可能になるものと考えられる．

6. 結論

本章では，本研究のまとめと今後の課題について述べる．

6.1 まとめ

本研究では，サイバー攻撃に対する“積極的防御”として，機械学習を活用することにより，ダークウェブから効率的にインテリジェンスを抽出することを目指した．さらに，ダークウェブ上にフォーラムが乱立している現状に着目し，フォーラムの特性を明らかにする手法を提案した．

実験により，ニューラルネットワークを応用した doc2vec は，機械学習における自然言語処理・特徴抽出の手法として高い性能を持つこと分かった．考察で述べたとおり，未知データの分類には今後も試行錯誤が必要であるが，実験で使用したデータ数に比して，SVM 及び MLP は，いずれも 90%以上という高い分類性能を示した．これは，doc2vec のベクトル化が投稿の特徴を正確に表現した証左である．

さらに，実験により，ダークウェブの投稿に対し，機械学習を用いることが有用であることが分かった．ダークウェブには，玉石混淆の投稿が存在しているため，大規模なデータに内在する規則性を導出する機械学習は，ダークウェブの分析に極めて相性が良いものと思料される．実際，実験で生成したモデルは，doc2vec との組み合わせにより高い汎化性能を示した．

以上のとおり，本研究で提案した手法は，いまだ完全なものとは言いがたいが，次節で述べる課題を克服することで「機械学習を用いたインテリジェンス抽出手法」として効果的であると考えられる．本研究の成果は，手法確立に至る方法論を示したことである．

6.2 今後の課題

今後の課題として，以下の 4 点を挙げる．

第 1 に，第 7 章の考察で述べたとおり，実験では，機械学習に用いるデータの少なさが課題となった．機械学習は，データの数が学習の精度に直結する．ダークウェブからデータを収集することは簡単ではないが，ツールを使用するなどしてより多くのデータを収集する必要がある．この点，Sixgill は，極めて優れたツールであり，時間的制約がなければより多くのデータを収集できたものと思われる．

第 2 に，第 7 章の考察で述べたとおり，ラベル付けを慎重に行う必要がある．機械学習の分類精度は，人間のラベル付けに依存する．本実験では，「マルウェア・オファーに関連する投稿」を正解データとしてラベル付けすることに神経を使ったが，不正解データにも同様の注意を払うべきであった．分類の目的は，正解データを抽出することではなく，あくまで正解データと不正解データを分類することにあるからである．したがって，今後は，不正解データにいかなるデータを入れるかにも関心を寄せる必要がある．

第 3 に，提案手法の目的は，フォーラムの特性を明らかにすることである．実験では，「マルウェア・オファーのプラットフォーム」として機能するフォーラムを抽出することを目指した．今後は，これだけでなく「ハッキング技術」や「クレジットカード情報の売買」，「サイバー攻撃の勧誘」のプラットフォームとして機能するフォーラムを抽出することで，ダークウェブに乱立する多くのフォーラムの特性を明らかにする必要がある．

第 4 に，提案手法は，「データの準備」，「自然言語処理・特徴抽出（doc2vec）」，「機械学習」及び「未知データの分類・分析」という 4 つのフェーズに分けられる．現状，これらのフェーズは個々に独立している．してがって，長期的な課題ではあるが，これらを一括して自動化できる仕組みを構築する必要がある．

謝辞

本研究において，専門家の見地から本研究に実用性を付与して下さったデロイトトーマツリスクサービス株式会社の皆様に感謝致します．

また，優れたツールである Sixgill を御厚意により無償で提供して下さったイスラエルの Sixgill 社の皆様にもお礼申し上げます．

参考文献

[1] AV-TEST, https://www.av-test.org/fileadmin/pdf/security_report/AV-TEST_Security_Report_2016-2017.pdf

[2] マーク・M・ローエンタール(2013)『インテリジェンス　機密から政策へ』慶應義塾大学出版会　2pp.

[3] デロイトトーマツグループ, https://www2.deloitte.com/jp/ja/pages/about-deloitte/articles/news-releases/nr20160524.html

[4] KELA GROUP, https://site.ke-la.com/cyber/jp/cyber-intelligence

[5] 株式会社テリロジー, http://www.terilogy.com/product/cyberthreatintelligence/index.html

[6] Kristin Finklea , “Dark Web”, Congressional Research Service, 2017

[7] Michael Chertoff, Toby Simon, “The Impact of the Dark Web on Internet Governance and Cyber Security”, Global Commission on Internet Governance, Paper Series: No. 6, February 2015

[8] Benjamin Brown, Akamai SIRT, “2016 State of the Dark Web“, https://www.akamai.com/cn/zh/multimedia/documents/state-of-the-internet/akamai-2016-state-of-the-dark-web.pdf, 2016

[9] Tomas Mikolov, “Distributed Representations of Sentences and Documents”, Proceedings of The 31st International Conference on Machine Learning(ICML), 2014

[10] Mikolov Tomas, Sutskever Ilya, Chen Kai, Corrado, Greg, and Dean, Jeffrey, “Distributed representations of phrases and their compositionality”, In Advances on Neural Information Processing Systems, 2013

[11] Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin, Vineet Mishra, Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, Paulo Shakarian, “Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence”, Cryptography and Security(cs.CR), 2016

[12] Sagar Samtani, Ryan Chinn, Hsinchun Chen, “Exploring Hacker Assets in Underground Forums”, IEEE Intelligence and Security Informatics, 2015

[13] Ahmed Abbasi, Weifeng Li, Victor Benjamin, Shiyu Hu, Hsinchun Chen, “Descriptive Analytics: Examinig Expert Hackers in Web Forums”, IEEE Intelligence and Security Informatics, 2014

[14] Victor Benjamin, Weifeng Li., Thomas Holt, Hsinchun Chen, “Exploring Threats and Vulnerabilities in Hacher Web: Forums, IRC and Carding Shops”, IEEE Intelligence and Security Informatics, 2015

[15] Mitch Macdonald, Richard Frank, Joseph Mei, Bryan Monk, “Identifying Digital Threats in a Hacker Web Forum” IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2015

[16] Sixgill, https://www.cybersixgill.com/

[17] genism, https://radimrehurek.com/gensim/

[18] scikit-learn, http://scikit-learn.org/stable/#