情報処理学会論文誌 Vol.62 No.9 1575–1585 (Sep. 2021)

専門家の知見に基づいた特徴量設計による IDPS シグネチャ重要度分類

川口 英俊[1,2,a),b)] 中谷 裕一[1,c)] 岡田 将吾[2,d)]

受付日 2020年12月26日, 採録日 2021年6月7日

概要：IDPS（Intrusion Detection and Prevention Systems）を実現するうえで，悪性通信のパターンファイルであるシグネチャの管理コストが増加する．本研究はそのコスト軽減を目的としており，機械学習によるシグネチャの自動分類モデルを提案・評価する．自動分類の精度向上のため，専門家の知見を参考に，シンボル特徴量（SF），キーワード特徴量（KF），WEB-MSG 特徴量（WMF）の 3 つを提案する．実験には，専門家が作成した (i) If/Then ルールに適合するデータセット，(ii) If/Then ルールで分類できないデータセットの 2 つを用いる．提案した特徴量が有効であることを，複数の機械学習分類モデルを用いた実験で示す．balanced-accuracy を計測した結果，SF と KF を用いた場合 (i) では 95.7%の性能を確認できたものの，(ii) では 59.6%という結果だった．一方で，提案する SF と WMF も組み合わせることで，(ii) においても 86.8%の精度を得ることができた．また，追加実験により分類に有効な特徴量を明らかにした．

キーワード：機械学習，特徴量設計，Snort，IDPS，シグネチャ，tf-idf

Classification of IDPS Signature Importance with Feature Engineering Based on Expert's Knowledge

Hidetoshi Kawaguchi[1,2,a),b)] Yuichi Nakatani[1,c)] Shogo Okada[2,d)]

Received: December 26, 2020, Accepted: June 7, 2021

Abstract: In order to realize IDPS (Intrusion Detection and Prevention Systems), the management cost of signatures, which are pattern files of malicious communication, increases. In this study, we propose and evaluate an automatic classification model of signatures using machine learning to reduce the cost. In order to improve the accuracy of automatic classification, we propose three types of features: symbol features (SF), keyword features (KF), and WEB-MSG features (WMF) based on the experts' knowledge. We use two datasets in our experiments: (i) data sets that can be classified by the expert-created If/Then rule, and (ii) data sets that do not match the If/Then rule. The effectiveness of the proposed features is shown in experiments using several machine learning classification models. Using the dataset (i), the balanced accuracy was 95.7% for the combined features of SF and KF. In (ii), the accuracy of the model trained with SF and WMF was 59.6%. On the other hand, by combining the proposed SF and WMF as well, we obtained an accuracy of 86.8% in (ii). Additional experiments revealed features that are useful for classification.

Keywords: machine-learning, feature-engineering, Snort, IDPS, signatures, tf-idf

1. はじめに

IDPS（Intrusion Detection and Prevention Systems）は情報通信システムを監視し，悪性通信を検知した際にロギング・通知・遮断等のアクションを行うシステムである．本稿では，IDPS シグネチャ（以下，シグネチャ）等の悪性通信のパターンファイルをもとに検知を行うタイプの IDPS に着目する [1]．シグネチャは，IDPS を開発・販売し

[1] NTT ネットワークサービスシステム研究所
Network Service Systems Laboratories, NTT, Musashino, Tokyo 180–8585, Japan
[2] 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology, Nomi, Ishikawa 923–1211, Japan
[a)] hidetoshi.kawaguchi.my@hco.ntt.co.jp
[b)] kawa.hide39@jaist.ac.jp
[c)] yuichi.nakatani.rd@hco.ntt.co.jp
[d)] okada-s@jaist.ac.jp

ている企業によって定期的に配布される．IDPS のユーザは，定期的に配布されるシグネチャを受け取った後に，シグネチャごとに悪性通信を検知した際の IDPS のアクションを設定する必要がある．

専門家はシグネチャごとにその重要度を判定し，その重要度をもとに IDPS のアクションを設定している．たとえば，シグネチャの重要度が高いのでれば “遮断” を，低いのであれば “ロギング” というアクションを設定することになる．この重要度の判定は，専門性と時間が必要であり，セキュリティ運用の業務として大きなコストと考えることができる．本稿では，専門家がシグネチャの重要度を判定することを，分類問題として取り扱うものとする．

本研究ではシグネチャの管理コスト軽減を目的に，シグネチャの重要度分類に有効な特徴量を設計し，機械学習による分類モデルを構築・評価する．有効な特徴量を設計するために，専門家に重要度の判断に関してヒアリングを行った．専門家は，以下の手順に従いシグネチャを半自動で分類している．まず，自作の If/Then ルールを適用することで分類を行う．この If/Then ルールは，シグネチャ内の要素のキーワードマッチングの組合せを条件として重要度ラベル，または分類不明という意味のラベルをシグネチャに付与する．その後，If/Then ルールで分類不明と判定されたシグネチャを手動で分類する．

本研究では単一の分類モデルですべてのシグネチャの分類を行うものとする．そのため，If/Then ルールに適合するシグネチャ，適合しないシグネチャに関わらず共通の特徴量ベクトルとして設計・表現する．上記の専門家の分類方法を参考に，(1) If/Then ルールの条件の対象となっている特徴量，(2) If/Then ルール内のキーワードをもとに得られる特徴量，(3) シグネチャ内のメッセージと外部参照情報から Web スクレイピングにより取得した言語特徴量の，3 つを設計する．

評価実験には，実際のシグネチャに専門家がラベルを付与して作成した 2 つの実データセットを用いる．1 つは If/Then ルールで分類可能なデータセットであり，もう一方は If/Then ルールで分類できないシグネチャで構成されたデータセットである．本研究で設計した特徴量が 2 つのデータセットの分類タスクに有効であることを，複数の機械学習分類モデルを用いた実験で示す．

本稿の貢献は以下のとおりである．

- 重要度ラベル付きの IDPS シグネチャのデータセットを新規に構築し，自動分類モデルを提案・評価する．
- IDPS シグネチャの特徴量を提案する．
- 専門家が手動でシグネチャを分類する場合，シグネチャ内の文字列情報や WWW 上の情報を重要視していることを，特徴量の分析により示す．

シグネチャの重要度は，IDPS が監視する情報通信システムに依存している．言い換えれば，同じ専門家が同じシグネチャの重要度を判断したとしても，IDPS が監視する情報通信システムによってシグネチャの重要度は異なる可能性がある．そのため，提案する分類モデルは，IDPS が監視する情報通信システムごとに構築する必要がある．したがって，複数の情報通信システムでのシグネチャ分類モデルの構築・評価，情報通信システム間での分類モデルの違いの分析等は今後の課題とし，本研究では扱わない．このため他の情報通信システムで収集したデータセットを用いた場合に同様の精度を保証することはできないものの，専門家によるデータセットの作成方法，特徴量の抽出方法，分類モデルの構築方法は IDPS が監視する情報通信システムに依存しない．

2 章では，本研究の位置づけを示すために，IDPS の検知力向上および運用負担軽減に関する研究について述べる．3 章では，本稿で対象とするデータセットや問題設定について述べる．4 章では，提案する特徴量設計について述べる．5 章では，実験により，提案した特徴量での機械学習分類モデルの性能を確認し，分類に有効な特徴量を分析する．6 章でまとめと今後の課題について述べる．

2. 関連研究

IDPS とその運用は，情報通信システムをサイバー攻撃から防衛するための重要な業務に位置づけられており，IDPS の検知力向上に関する研究（2.1 節）や，IDPS の運用負担を軽減するための研究（2.2 節）はさかんに行われている．これらの研究を概観し，本研究の位置付けを述べる．

2.1 IDPS の検知力向上に関する研究

IDPS は，悪性通信のパターンをあらかじめセットしておき，それに一致する通信に対してロギング・通知・遮断等のアクションを行う．悪性通信のパターンの表現形式は，シグネチャと機械学習による分類モデルの 2 種類がある．IDPS のベンダがサイバー攻撃を分析しながらシグネチャを手動で作成し，提供している．機械学習よる分類モデルも，一般的にはベンダが開発を担っている．

シグネチャの作成はベンダにとっても負担が大きいことから，その負担を軽減するために，自動で生成するための研究が行われてきた．Shahriar らは過去のシグネチャをもとに遺伝的アルゴリズムでシグネチャを自動生成する手法を提案している [2]．その他，Fallahi らは決定木を [3]，Lee らは LDA（Latent Dirichlet Allocation）をシグネチャの自動生成に応用している [4]．

機械学習による正常な通信と悪性通信の分類モデルに関する研究は数多く行われている [1]．この場合，機械学習による分類モデルは，悪性通信の特徴量を入力として，悪性通信か否か，もしくは悪性通信の種類を多クラス分類として出力する．SVM（Support Vector Machines），決定木，バギング，人工ニューラルネットワーク等の手法が適用さ

れている [5], [6], [7], [8]．この分野の研究は統一的なベンチマークテスト（NSL-KDD，UNSW-NB15，TUIDS 等）も整備されており [9], [10], [11]，数多くの研究成果が報告されている．

以上の研究は，IDPS の検知力向上に寄与すると考えられる．一方，IDPS の運用負担を軽減するための研究も行われており，本研究はその研究分野に属する．次節で，その研究分野について述べる．

2.2 IDPS 運用の負担軽減に関する研究

IDPS の検知力向上が重要である一方，セキュリティ運用の現場では IDPS を効率的に運用することも重要である．IDPS の運用で特に負担となるのは，誤検知によるアラートへの対応とシグネチャの管理である．

通常，多くのアラートは IDPS の誤検知によるものであり，IDPS のユーザはその対応に日々追われている．この負担を軽減するために，IDPS からのアラートを分析し，アラートそのものを削減するための研究が行われている．Alsubhi らは，アラートの優先順位を推定するファジイ理論システムを提案している [12]．Pietraszek はアラートの誤報を削減するための，機械学習を組み込んだシステムを提案している [13]．

日々作成され続けるシグネチャを適切に整理・取捨選択するための研究も行われている．Stakhanova らは，矛盾するシグネチャどうしを発見するための分析モデルを提案している [14]．その分析モデルでは，シグネチャは非決定性オートマトンで表現され，オートマトンの等価性をもとにシグネチャの重複を検知している．同様の目的で，Massicotte らも集合論とオートマトンの理論をベースにした別のアプローチを提案している [15]．

我々の研究は，シグネチャの管理に関する研究に分類される．1 章で述べたように，シグネチャは 1 件ごとに重要度を判定する必要がある．しかしながら，我々の知る限りでは，この判定を自動化するための研究は行われていなかった．シグネチャの重要度判定を自動化するためには，専門家の暗黙知や思考パターンを適切にモデリングする必要がある．そこで本稿では，そのモデリングを機械学習を使って実現し，評価・分析を行う．

3. 問題設定

3.1 データセット

本稿では，実際にセキュリティ運用に従事している専門家がラベル付けしたシグネチャを対象に，実験・分析を行う．その専門家は実際のセキュリティ運用で使用するための If/Then ルールを設計した．If/Then ルールは，シグネチャ内の要素のキーワードマッチングを条件として重要度ラベル（“low”，“medium”，“high”）または分類不明（“unknown”）ラベルを返す．これらの重要度は，そのシグネチャがマッチしたときの IDPS のシグネチャの挙動と対応している．“low” の場合は，専門家への通知も通信の遮断も行わない．“medium” の場合は，通知のみを行う．“high” の場合は，通知に加えて通信の遮断を行う．まず専門家は作成した If/Then ルールで自動的に分類した．次に，If/Then ルールに “unknown” と分類されたシグネチャを，専門家は手動でラベル付けした．これにより，If/Then ルールで分類されたシグネチャと手動で分類されたシグネチャの 2 つのデータセットが作成された．前者を AAD（Automatically Annotated Dataset），後者を MAD（Manually Annotated Dataset）と呼称するものとする．

表 1 データセットの情報
Table 1 Summary of datasets.

Dataset	Priority			Total
	low	medium	high	
AAD	3,936	93	436	4,465
MAD	1,119	122	59	1,300

IDPS は，実際にインターネット上でサービスを提供しているサーバの通信を監視することを目的に運用されている．このサービスは，専門家の所属組織と契約したユーザのみが利用することができるものである．本研究のデータセット（AAD と MAD）は，この IDPS に用いるために作成された実データセットである．データセットを構成するシグネチャは，2016 年 12 月から 2018 年 5 月の約 1 年 6 カ月間に，IDPS を開発・販売している企業から配布されたものである．

表 1 は専門家が作成した AAD と MAD のサンプル数とその内訳を示している．シグネチャは “low”，“medium”，“high” の 3 種類のいずれかの重要度ラベルが付与されている．専門家はこれらの重要度に基づいて，そのシグネチャとマッチした通信に対する IDPS のアクションを設定している．

3.2 シグネチャの構造

対象とするシグネチャは Snort [*1] という IDPS のセキュリティエンジンに対応した記法で記述される．記法を説明するために，具体例を**図 1** に示す[*2]．先頭の “alert” に位置する文字列は，そのシグネチャで検知した際の IDPS のアクションを示す．アクションは専門家が重要度に基づいて設定するため，分類モデルに入力することはできない．そのため，アクション以降の文字列から特徴量を抽出する．

“tcp $HOME_NET any -> $EXTERNAL_NET any” はスペース区切りで 5-tuple の情報が記載されている．5-tuple とは，IP パケットのヘッダ内に記載されている 5 つの情

[*1] https://snort.org/（2020 年 6 月 12 日）
[*2] Snort の公式ページで配布されているものから引用している．https://www.snort.org/downloads/#rule-downloads（2020 年 6 月 12 日）

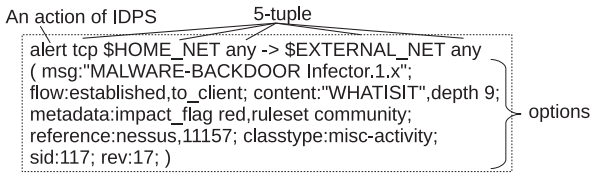


図 1 シグネチャの具体例
Fig. 1 A specific example of signatures.

報のセットのことであり，通信プロトコル，送信元 IP アドレス，送信元ポート番号，宛先 IP アドレス，宛先ポート番号の 5 つで構成されている．中心付近にある -> は通信の方向を表している．“tcp”，“$HOME_NET”，“any”，“$EXTERNAL_NET”，“any” は順に通信プロトコル，送信元 IP アドレス，送信元ポート番号，宛先 IP アドレス，宛先ポート番号を表している．以上の情報がシグネチャに必須の情報であり，以降の括弧内はシグネチャ作成者が任意で記述可能なオプションである．オプションは基本的には key-value 形式で表されるが，以下の特徴がある．

- key と value のひも付けは ‘:’ で行う．
- key-value の区切りに ‘;’ を用いる．
- 同じ key を複数個許容できる key-value がある．
- “nocase” 等，key の存在しない値もある．

オプションには多くの要素があるが，本評価では msg（message の略称），metadata，reference，classtype という 4 つの key-value に着目する．これらの要素は，専門家が作成した If/Then ルールが参照するすべての要素であり，If/Then ルールはこれらのオプションの要素と 5-tuple のみから分類を行う．

msg は，シグネチャにマッチしたときにログやアラートに記載する文字列である．図 1 の “MALWARE-BACKDOOR Infector.1.x” が msg の一例である．

metadata は，シグネチャ全体のオプションと同じく key-value 形式で自由に情報を記述できる要素である．具体例の “impact_flag red,ruleset community” が該当する．key と value のひも付けは半角スペースで行われ，key-value のセットの区切りには ‘,’ が使われる．

reference は，外部の情報を参照するための情報が記載されている．具体例では “nessus,11157” と記載されているが，これは nessus という外部システムが管理する ID11157 の情報を指している．reference の記述方法は 2 種類ある．1 つ目は，製品の脆弱性に関する情報リスト（以下，脆弱性リスト）の名前と ID が記載される．脆弱性リストは nessus のほかには，共通脆弱性識別子（Common Vulnerabilities and Exposures，CVE）や Bugtraq がある．ID が 1999-0067 の CVE の場合は “cve,CVE-1999-0067” と記載され，ID が 629 の Bugtraq は “bugtraq,629” と記載される．2 つ目は，情報へのアクセス先として URL が直接記載される．たとえば “url,www.spywareguide.com/product_show.php?id=973” となる．

classtype はそのシグネチャが示す悪性通信のグループを示している．具体例の “misc-activity” が該当する．このグループは，専門家が判断する重要度とは異なる．

3.3 If/Then ルールについて

If/Then ルールは，キーワードマッチングの組合せを条件としていずれかの重要度ラベルまたは “unknown” ラベルを付与する．このキーワードマッチングとは，ある単語を含むか否かの判定のことである．キーワードマッチングは 5-tuple，msg，metadata，reference，classtype を対象とする．metadata については，特定の key と value の組合せを 1 つのキーワードとしてマッチングを行っている．msg については，ある単語が msg 中に存在しているか否かで判定しており，単語の位置は考慮していない．reference は，特定のシステムを参照しているか否かを判定しており，ID は条件に用いていない．5-tuple は通信プロトコル，送信元 IP アドレス，送信元ポート番号，宛先 IP アドレス，宛先ポート番号を個別に抽出し，それぞれで判定を行っている．classtype は，単一の記号で表現されているため，特別な前処理は施していない．

実際に If/Then ルールを作る際に抽出されたキーワード数は，5-tuple が 133 個，metadata が 2 個，msg が 56 個，reference が 1 個，classtype が 6 個となった．これらへのキーワードマッチングを基本構成要素として，論理積や論理和を組み合わせた条件 61 個が If/Then ルールに含まれている．複数キーワードとの一致を条件とする判定も存在するため，条件数はキーワード数よりも少なくなっている．キーワードの具体例として，重要度が “high” となる条件には classification の “trojan-activity” や msg の “MALWARE-TOOLS” 等がある．重要度が “medium” の場合は classification の “network-scan” や msg の “BLACKLIST” 等がある．重要度が “low” の場合には，5-tuple の送信元 IP アドレスの “$EXTERNAL_NET” や，msg の “MALWARE-CNC” 等がある．

If/Then ルールは，条件に適合したシグネチャのみに重要度ラベルを付与する．いずれの条件にも適合しなかったシグネチャは，“unknown” ラベルが付与される．条件には優先順位が設定されており，複数の条件に適合する場合は，優先順位が高い条件の重要度ラベルが付与される．

専門家は，可能な限り重要度ラベルを付与できるように If/Then ルールを作成した．If/Then ルールのみでシグネチャを分類できれば理想的だが，実際には 1,300 件のシグネチャには If/Then ルールではラベルを付与することができなかった（表 1）．これらのシグネチャを分類するためには，専門家の高度な専門性や知識が必要である．専門家による手動ラベル付けの思考をモデル化することは容易ではないが，機械学習によりその解決を目指す．

4. 特徴量設計

本章では，シンボル特徴量（symbol features，SF），キーワード特徴量（keyword features，KF），WEB-MSG 特徴量（web information and message features，WMF）の 3 つの特徴量を設計する．これらの特徴量への変換手順と関係を**図 2** に示す．SF と KF は If/Then ルールを，WMF は専門家へのヒアリング結果を参考に設計を行っている．SF は 5-tuple，metadata，classtype の 3 つを，KF と WMF は msg と reference を対象としている．

4.1 シンボル特徴量

SF は，5-tuple，metadata，classtype を対象に，それぞれを One-hot エンコーディングで特徴量として抽出する．処理手順を図 2 の左に示している．One-hot エンコーディングは，名義特徴量を数値ベクトルに変換する手法である．たとえば A，B，C の 3 種類の記号がある場合，それぞれ $[1,0,0]$，$[0,1,0]$，$[0,0,1]$ と特徴量に変換される．

classtype は One-hot エンコーディングでそのまま特徴量に変換可能である．しかし，5-tuple と metadata については前処理が必要がある．5-tuple は，通信プロトコル，送信元 IP アドレス，送信元ポート番号，宛先 IP アドレス，宛先ポート番号の 5 つに分解する．その後，それぞれに One-hot エンコーディングで数値ベクトルに変換する．metadata については，まずは key-value を 1 つの記号（key-value 記号）と見なしてすべて抜き出す．その後，抜き出した key-value 記号を並べ替えて文字列として結合し，1 つの記号と見なす．この並べ替えは，metadata 内で現れた key-value の順序の影響をなくすために行っている．原理上は，key-value の組合せは莫大な種類が存在するが，実際に AAD と MAD に現れている種類の数は少数である．

4.2 キーワード特徴量

KF は，If/Then ルール内の msg と reference を対象としたキーワードマッチングを参考に設計されている．KF の処理手順を図 2 の右上に示している．If/Then ルール内に存在するキーワードが，msg と reference に現れているかどうかを抽出する．キーワードの抽出後は，5-tuple，metadata，classtype と同様に，One-hot エンコーディングで数値ベクトルに変換する．

msg については，If/Then ルールの条件に使われている単語のリストを作成しておく．その単語リストに合致する単語だけを msg から抜き出し，複数個ある場合は一定の規則で並べ替えてから結合し，1 つの記号と見なす．単語リストに合致する単語が 1 つも存在しない場合は，そのことを意味するダミー記号として抽出する．その後，One-hot

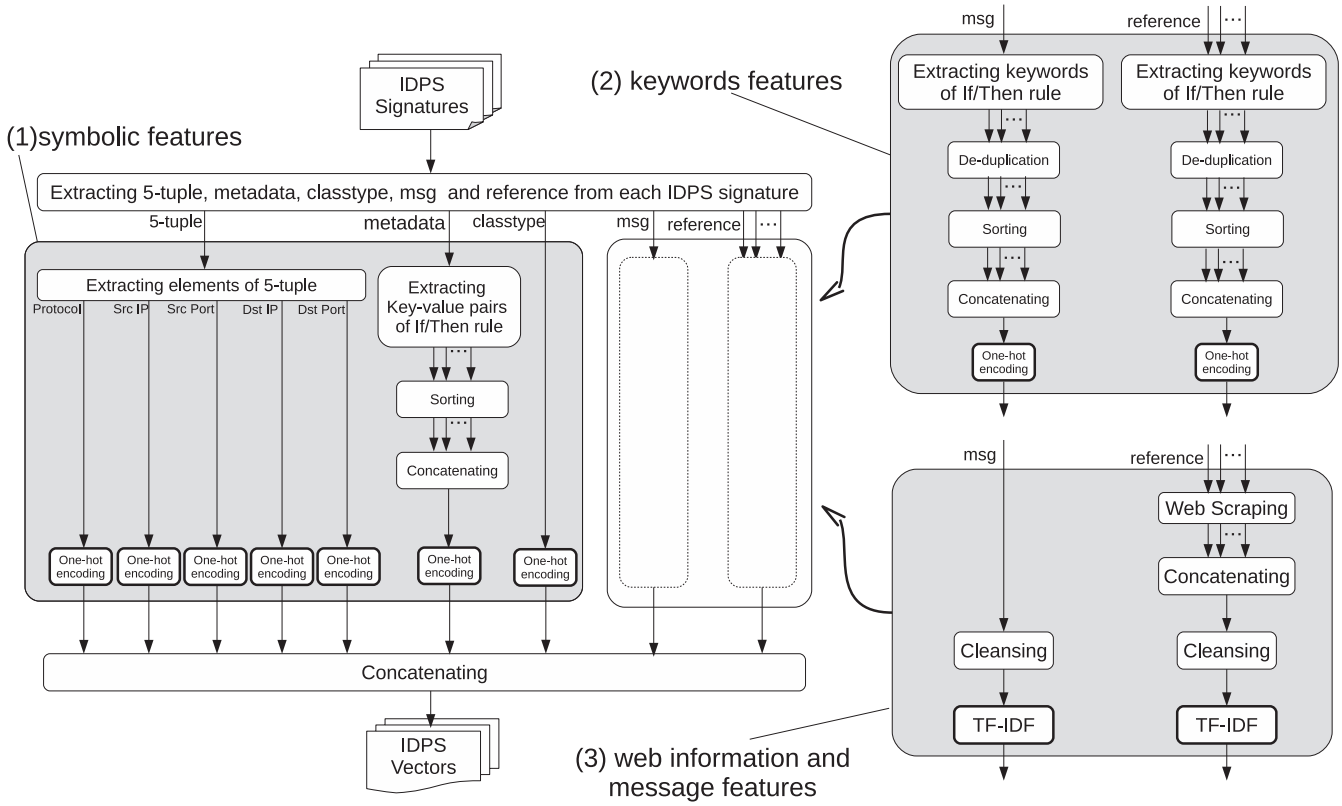


図 2 提案する特徴量：(1) シンボル特徴量，(2) キーワード特徴量，(3) WEB-MSG 特徴量
Fig. 2 Proposed features: (1) symbolic features, (2) keyword features, (3) web information and message features.

エンコーディングで記号を数値ベクトルに変換する．

reference も msg と同じ処理で数値ベクトルに変換される．msg の場合と異なる点は，msg では単語のリストを作成したが，代わりに If/Then ルールの条件に使われているシステム名のリストを作成しておく．そのシステム名のリストを使い，以降は msg の場合と同じ処理を行う．

4.3 WEB-MSG 特徴量

実際には If/Then ルールで分類できないシグネチャが多数存在するため，上述した特徴量に加え，新たな判断基準を追加し，特徴量を拡張する必要がある．専門家が手動で分類する場合の知見から新たな特徴量を設計するため，専門家にヒアリングを行った．

専門家は，主に msg や reference の指す外部情報を基に，運用対象である情報通信システムの構成情報や自身の経験・知識とあわせて統合的に重要度を判断している．具体的にはまず，msg と reference の示す外部情報からそのシグネチャが対象としている悪性通信の種別や細かな特性を把握する．そして，それらの種別や特性が情報通信システムに与える悪影響の度合いや，過検知・誤遮断のリスクを加味し，重要度を決定する．たとえば，悪性通信の種別が，データベースを攻撃する SQL インジェクションであり，情報通信システム内でデータベースが管理されている場合は重要度を “medium” か “high” とする可能性が高くなる．

以上のヒアリング結果から，msg と reference に重要な判断材料が含まれていると考え，それらの情報を拡充した特徴量を提案する．その拡充のために，自然言語分析で頻繁に用いられている tf-idf と，Web スクレイピングを用いるものとする．これらの手法を組み合わせ，msg と reference の情報を拡充する．特徴量の抽出手順を図 2 の右下に示す．

reference については，その外部情報の参照先から Web スクレイピングを行い情報を取得する．reference は脆弱性リスト（CVE や Bugtraq 等）の名前とその ID のセットもしくは URL で記載されているため，シグネチャに関連する情報を一意に特定できる．たとえば CVE を参照する場合は，NVD（National Vulnerability Database）[*3]や RedHat の CVE Database [*4]等の Web システムから ID で検索することで情報を取得可能である．シグネチャに関連する情報の具体例として，シグネチャが示す悪性通信が対象としているソフトウェアやそのバージョン情報が該当する．分類モデルの構築者は，基本的には reference が参照する情報を公開している Web システムごとに Web スクレイピングの処理手順を記述する必要がある．あらゆる Web システムに対応することは困難だが，参照頻度の高い Web システムに対象を絞れば現実的に記述可能である．以降で reference と指すものはこの Web スクレイピングで取得した情報のことを指すものとする．

msg，reference それぞれを文書と見なしてクレンジングを行い，個別に tf-idf で特徴量変換を行う．tf-idf とは，文書群を数値ベクトル群に変換する，自然言語処理に用いられてきた実績のある手法である [16], [17], [18]．tf-idf は文書を tf と idf を乗じた数値で構成される数値ベクトルに変換する．ここで tf は文書内のその単語の出現数を表し，idf は逆文書頻度という単語の珍しさ表している．

5. 実験

提案した特徴量の性能を検証するために，AAD と MAD を対象として実験・分析を行う．実用上は，If/Then ルールで分類が可能な AAD を機械学習モデルで分類する必要はない．しかし，本研究の目的はシグネチャの分類モデルの構築であり，提案する特徴量と機械学習で If/Then ルールをどの程度模擬できているかも確認するために，AAD も対象に実験を行う．

5.1 データセットから特徴量ベクトルへの変換

SF と KF を連結した特徴量を If/Then ルール特徴量（If/Then rule features，ITRF），SF と WMF を連結した特徴量を手動分類特徴量（manual classification features，MCF）とする．ITRF は If/Then ルールを参考に，MCF は手動での分類を参考にした特徴量である．

MCF に変換する際に，AAD と MAD 両方のシグネチャ 1 件ごとに，Web スクレイピングで情報の拡張を行う．本実験では，シグネチャが示す悪性通信の対象となるソフトウェアやバージョン情報を Web スクレイピングで取得する．シグネチャに記載されている CVE，Bugtraq，URL の順に，対象となるソフトウェア情報およびバージョン情報を示すテキストの取得を試行し，取得に成功した時点で当該シグネチャでの Web スクレイピングは終了する．CVE から取得する場合は，NVD，RedHat の CVE Database の順に検索を行う．Bugtraq から取得する場合は，SecurityFocus [*5] から検索を行う．URL から取得する場合は，その参照先が，Talosintelligence の Vulnerability Report [*6]，Adobe Security Bulletin [*7]，Exploit Database [*8]の場合に取得を試みる．以上の手順で，AAD は 4,465 件中 2,807 件，MAD は 1,300 件中 1,024 件の情報取得に成功した．

MCF の WMF への変換の際の文字列へのクレンジングとして，以下の処理を行っている．まず，アルファベット，数値とアンダーバーのみを用い，それ以外の記号は半角スペースに置き換える．次に，英語でストップワードと呼ば

[*3] https://nvd.nist.gov/vuln（2021 年 4 月 16 日）
[*4] https://access.redhat.com/security/security-updates/#/cve（2021 年 4 月 16 日）
[*5] https://www.securityfocus.com/（2021 年 4 月 16 日）
[*6] https://talosintelligence.com/vulnerability_reports（2021 年 4 月 16 日）
[*7] https://helpx.adobe.com/security.html（2021 年 4 月 16 日）
[*8] https://www.exploit-db.com/（2021 年 4 月 16 日）

れる単語 [19] や全シグネチャ内で 1 回しか現れなかった単語は削除する．

クレンジングされた msg と reference に対して，それぞれ個別に tf-idf で特徴量ベクトルに変換する．テキスト（シグネチャ内の msg もしくは reference）を示す識別子を d，単語を示す識別子を t とした場合の，tf-idf は以下のとおりである[*9]．

$$tf\text{-}idf(t,d) = tf(t,d) \cdot idf(t) \tag{1}$$

ここで，$tf(t,d)$ はテキスト d に現れる単語 t の出現数（0 以上の整数）を表す．$idf(t)$ は以下のとおりである．

$$idf(t) = \log \frac{N_L + 1}{df_L(t) + 1} + 1 \tag{2}$$

N_L は，訓練データとテストデータを合わせた全テキスト数ではなく，訓練データのみのテキスト数である．$df_L(t)$ は，N_L 個の訓練データにおける，単語 t の現れたテキスト数である．つまり，テストデータを tf-idf で特徴量ベクトルに変換する際の idf は，訓練データ内で算出された idf を用いていることになる．tf-idf は，すべての単語をユニグラムとして扱う．また，L2 正規化を行い，最小値 0，最大値 1 で min-max スケーリングを行う．

5.2 機械学習モデル

線形 SVM（Linear-SVM），多層ニューラルネットワーク（Multilayer Perceptron），決定木（Decision Tree），ランダムフォレスト（Random Forest），ナイーブベイズ（Naive Bayes）用いて，評価実験を行う．

Linear-SVM の正則化パラメータ C を 1.0 として学習する．また，One-vs-Rest で多クラス分類に対応する．

Multilayer Perceptron は，ノード数 100 の中間層を持つ 3 層構造として，誤差逆伝播法で学習する．すべてのノードの活性化関数は ReLU（ランプ関数）とする．L2 正則化で過学習を抑える．正則化パラメータは 0.0001 とする．最適化手法には Adam（Adaptive moment estimation）を提案論文 [20] に記載されているデフォルトパラメータ（$\alpha = 0.0001$，$\beta_1 = 0.9$，$\beta_2 = 0.99$，$\epsilon = 10^{-8}$）で用いる．

Decision Tree はジニ不純度を指標とした CART（Classification and Regression Tree）法で学習する．すべての端点（葉）に存在するサンプル数もしくはクラス数が 1 になるまで学習を行う．

Random Forest は Decision Tree と同様の方法で学習する決定木 10 個で構成する．それぞれの決定木は，特徴量ベクトルの次元数を m として $\lfloor \sqrt{m} \rfloor$ 個の特徴量をランダムに選択して学習する．$\lfloor \cdot \rfloor$ は床関数を表し，入力となる数値以下の最大の整数と定義される．

[*9] Python 言語の機械学習ライブラリである，scikit-learn（ver 0.21.2）の TfidfVectorizer クラスのデフォルト設定に従う．

ナイーブベイズでは，入力となる変数に正規分布を仮定して実装する．

5.3 実験結果

上述したデータセット（AAD と MAD），特徴量ベクトルへの変換方法（ITRF と MCF），機械学習モデル（Linear-SVM, Multilayer Perceptron, Decision Tree, Random Forest，Naive Bayes）の全組合せで，層化 10 分割交差検証（Stratified 10-fold cross validation）を行う．各 fold で，訓練データを 5.1 節に記載した方法で特徴量ベクトルに変換する．テストデータに関しても，同様に特徴量ベクトルに変換する．したがって，各 fold 間で特徴量ベクトルの次元数は異なる．特徴量ベクトルの次元数の平均と標準偏差を**表 2** に示す．AAD と MAD それぞれについて，ITRF と MCF の次元数を，本稿で着目したシグネチャの 5 つの要素（5tuple，metadata，classtype，msg，reference）ごとに記載している．左側の値が次元数の平均であり，括弧内の ± がついた値は標準偏差を表している．各数値は小数 2 位で四捨五入した値である．クラス間のサンプル数が不均衡であるため，交差検証の fold ごとに訓練データに対してオーバサンプリングを行う．3 クラスのうち，サンプル数の少ない 2 クラスのサンプルを，SMOTE [21] で最も多いクラスのサンプル数と同数まで増やす．サンプルを生成する際の基準サンプルからの近傍数は 5 とする．その後，機

表 2 実験における特徴量ベクトルの次元数
Table 2 Dimension of feature vector in the experiment.

Dataset	Features[†]		#. Dimensions
AAD	ITRF	5tuple	236.8 (±3.0)
		metadata	427.3 (±4.8)
		classtype	15.7 (±0.7)
		msg	21.9 (±0.3)
		reference	2.0 (±0.0)
	MCF	5tuple	236.8 (±3.0)
		metadata	427.3 (±4.8)
		classtype	15.7 (±0.7)
		msg	1,760.8 (±10.3)
		reference	2,081.8 (±75.0)
MAD	ITRF	5tuple	96.8 (±1.8)
		metadata	383.1 (±6.2)
		classtype	9.9 (±0.3)
		msg	6.0 (±0.0)
		reference	2.0 (±0.0)
	MCF	5tuple	96.8 (±1.8)
		metadata	383.1 (±6.2)
		classtype	9.9 (±0.3)
		msg	803.8 (±9.1)
		reference	563.2 (±27.1)

[†] ITRF がシンボル特徴量（SF）とキーワード特徴量（KF），MCF は SF と WEB-MSG 特徴量（WMF）の連結で構成されている．

表 3 提案した特徴量の性能評価
Table 3 Performance evaluation of the proposed features. ITRF is SF and KF, MCF is SF and WMF.

Dataset	Features †	Linear-SVM	Multilayer Perceptron	Decision Tree	Random Forest	Naive Bayes
AAD	ITRF	0.957 (±0.024)	0.957 (±0.020)	0.954 (±0.035)	0.929 (±0.039)	0.764 (±0.052)
	MCF	0.969 (±0.026)	0.964 (±0.024)	0.963 (±0.030)	0.926 (±0.031)	0.883 (±0.052)
MAD	ITRF	0.564 (±0.070)	0.590 (±0.067)	0.596 (±0.071)	0.573 (±0.101)	0.452 (±0.085)
	MCF	0.868 (±0.063)	0.865 (±0.070)	0.867 (±0.070)	0.824 (±0.044)	0.842 (±0.076)

† ITRF がシンボル特徴量（SF）とキーワード特徴量（KF），MCF は SF と WEB-MSG 特徴量（WMF）の連結で構成されている．

械学習で分類モデルを構築した．

層化 10 分割交差検証の各 fold で，構築済みの分類モデルでテストデータの分類を行い，balanced-accuracy を計測し，10 回分の fold 間で平均値と標準偏差を算出した．その結果を**表 3** に示している．Dataset 列は実験に用いたデータセットを表している．Features 列はデータセットを特徴量ベクトルに変換した方法を記している．各機械学習モデル名の列が，balanced-accuracy に関する値を示している．左側の値が balanced-accuracy の平均であり，括弧内の ± がついた値は標準偏差を表している．各数値は小数第 4 位で四捨五入した値である．

AAD に対する実験結果から，ITRF は十分に If/Then ルールに近い性能を発揮していることが分かる．一方で，MAD に対して ITRF では，AAD と比較して精度が大きく低下していることが分かる．これは If/Then ルールにマッチしていないデータセットである MAD を，If/Then ルールを参考に設計した特徴量で構成される ITRF では分類が難しいという結果を示している．

続いて，ITRF と MCF を比較するために，MAD を対象とした実験結果を確認する．すべての機械学習モデルで MCF のほうが性能が大きく上回っていることを確認できる．Linear-SVM は 0.304，Multilayer Perceptron は 0.275，Decision Tree は 0.271，Random Forest は 0.251，Naive Bayes は 0.390 といずれも最低 0.251 の性能向上を確認できる．MCF を用いた場合の性能の最高値は Linear-SVM の 0.868，最低値は Random Forest の 0.824 である．MCF は機械学習モデルによらず，最低でも 0.824 の性能が得られていることが分かる．ITRF と MCF の差分は，msg と reference に対する特徴量ベクトルの変換方法が KF か WMF かの違いのみである．そのため，いずれの機械学習モデルでも，WMF により最低でも 0.251 以上の精度改善が確認された．これらのことから，MCF に含まれる WMF は，専門家が手動で分類する場合の特性をよくとらえていると考えられる．

5.4 有効な特徴量の分析

本稿では，専門家へのヒアリング結果から msg と reference が重要であると仮定して WMF を設計した．その仮定の妥当性を確認するために，より詳細な実験を通じて有効な特徴量の分析を行う．上述した実験の条件と MCF で，MAD に対してシグネチャ内の要素である 5-tuple，metadata，classtype，msg，reference の全組合せ 31 通りで同様の実験を行った．その結果を**表 4** に示す．表記の都合上，5-tuple は 5t，msg は ms，metadata は mt，reference は rf，classtype は cl と省略して表記している．機械学習モデルごとに，balanced-accuracy の平均が最も高い値は太字で表記している．下線は使用した要素数ごとの，その機械学習モデルの中で最も高い値を示している．

全体として，msg と reference が性能向上に大きく貢献していることを確認できる．すべての機械学習モデルにおいて，msg と reference を含んでいる場合に最も性能が高くなっている．特徴量に変換した要素数ごとに比較を行っても，msg や reference を含んでいるものの性能が高い．5 つの要素単体という条件で比較すると，Naive Bayes 以外の機械学習モデルは，msg，reference，classtype，metadata，5tuple の順に性能が良い．Naive Bayes についても，上位 3 つは msg，reference，classtype という順であり，これは他の機械学習モデルと同様である．

これらのことから，msg と reference が重要であるという仮定が妥当であったと結論付ける．msg と reference は自然言語的な要素であり，これらの情報が支配的であるとするならば，自然言語処理の手法・技術を適用できる可能性がある．自然言語処理はディープラーニングの台頭により急速に進歩している分野の 1 つであり，その期待は大きい．たとえば，大規模な言語コーパスから学習した BERT（Bidirectional Encoder Representations from Transformers）モデル [22] の応用等が考えられる．BERT は多くのタスクに応用され優れた結果を残しており [23], [24], [25]，シグネチャの分類にも適用できる可能性は十分にある．

6. おわりに

本稿では，Snort に対応したシグネチャを機械学習モデルで分類するための特徴量である SF，KF，WMF を提案し，実験によりそれらの有効性を示した．SF と KF は If/Then ルールを参考に設計され，WMF は専門家へのヒアリングの結果を参考に設計された．WMF には，tf-idf や Web スクレイピングを組み合わせてその情報量を拡充するというアイデアを用いた．AAD と MAD という実データセッ

表 4 MAD に MCF を使用した場合の詳細な性能評価 †
Table 4 Detailed performance evaluation of the use of MCF in MAD.

Elements ‡	Linear-SVM	Multilayer Perceptron	Decision Tree	Random Forest	Naive Bayes
5t	0.469 (±0.079)	0.447 (±0.076)	0.431 (±0.060)	0.452 (±0.062)	0.371 (±0.066)
mt	0.499 (±0.053)	0.500 (±0.054)	0.509 (±0.060)	0.514 (±0.058)	0.346 (±0.041)
cl	0.575 (±0.079)	0.569 (±0.076)	0.575 (±0.079)	0.566 (±0.080)	0.437 (±0.052)
ms	0.850 (±0.048)	0.874 (±0.057)	0.851 (±0.083)	0.833 (±0.071)	0.830 (±0.077)
rf	0.808 (±0.075)	0.827 (±0.060)	0.806 (±0.078)	0.807 (±0.080)	0.767 (±0.045)
5t,mt	0.528 (±0.052)	0.552 (±0.062)	0.507 (±0.078)	0.499 (±0.062)	0.364 (±0.065)
5t,cl	0.611 (±0.084)	0.597 (±0.084)	0.576 (±0.106)	0.581 (±0.095)	0.443 (±0.093)
5t,ms	0.868 (±0.047)	0.852 (±0.059)	0.847 (±0.069)	0.763 (±0.061)	0.829 (±0.079)
5t,rf	0.864 (±0.070)	0.862 (±0.062)	0.817 (±0.064)	0.821 (±0.053)	0.773 (±0.049)
mt,cl	0.557 (±0.054)	0.575 (±0.080)	0.566 (±0.068)	0.542 (±0.079)	0.424 (±0.072)
mt,ms	0.782 (±0.061)	0.807 (±0.069)	0.834 (±0.066)	0.808 (±0.076)	0.829 (±0.072)
mt,rf	0.837 (±0.051)	0.840 (±0.076)	0.846 (±0.058)	0.843 (±0.054)	0.755 (±0.081)
cl,ms	0.843 (±0.049)	0.862 (±0.058)	0.843 (±0.061)	0.832 (±0.054)	0.829 (±0.082)
cl,rf	0.862 (±0.067)	0.847 (±0.072)	0.858 (±0.074)	0.855 (±0.079)	0.772 (±0.053)
ms,rf	**0.889 (±0.060)**	0.892 (±0.050)	0.875 (±0.070)	**0.884 (±0.070)**	**0.856 (±0.068)**
5t,mt,cl	0.566 (±0.068)	0.558 (±0.085)	0.547 (±0.108)	0.527 (±0.079)	0.436 (±0.090)
5t,mt,ms	0.798 (±0.068)	0.814 (±0.061)	0.817 (±0.083)	0.765 (±0.080)	0.827 (±0.075)
5t,mt,rf	0.835 (±0.088)	0.849 (±0.076)	0.835 (±0.064)	0.814 (±0.049)	0.736 (±0.093)
5t,cl,ms	0.853 (±0.050)	0.841 (±0.064)	0.823 (±0.069)	0.770 (±0.055)	0.827 (±0.083)
5t,cl,rf	0.866 (±0.069)	0.850 (±0.064)	0.806 (±0.062)	0.826 (±0.062)	0.778 (±0.067)
5t,ms,rf	**0.889 (±0.062)**	**0.895 (±0.063)**	0.867 (±0.074)	0.835 (±0.057)	0.854 (±0.068)
mt,cl,ms	0.808 (±0.058)	0.811 (±0.057)	0.831 (±0.066)	0.817 (±0.066)	0.827 (±0.079)
mt,cl,rf	0.844 (±0.048)	0.859 (±0.057)	0.846 (±0.078)	0.828 (±0.077)	0.756 (±0.074)
mt,ms,rf	0.871 (±0.068)	0.868 (±0.063)	0.868 (±0.077)	0.844 (±0.067)	0.847 (±0.069)
cl,ms,rf	0.879 (±0.055)	0.893 (±0.052)	0.867 (±0.071)	0.873 (±0.075)	0.851 (±0.078)
5t,mt,cl,ms	0.816 (±0.061)	0.810 (±0.059)	0.820 (±0.073)	0.718 (±0.057)	0.824 (±0.080)
5t,mt,cl,rf	0.839 (±0.079)	0.856 (±0.077)	0.831 (±0.082)	0.812 (±0.068)	0.739 (±0.090)
5t,mt,ms,rf	0.877 (±0.063)	0.868 (±0.064)	0.864 (±0.091)	0.840 (±0.056)	0.845 (±0.071)
5t,cl,ms,rf	0.888 (±0.053)	0.876 (±0.065)	0.875 (±0.069)	0.858 (±0.065)	0.848 (±0.078)
mt,cl,ms,rf	0.865 (±0.063)	0.873 (±0.066)	**0.894 (±0.047)**	0.847 (±0.062)	0.845 (±0.075)
5t,mt,cl,ms,rf	0.868 (±0.063)	0.865 (±0.070)	0.867 (±0.070)	0.824 (±0.044)	0.842 (±0.076)

† 下線は features の要素数ごとの，機械学習モデルの中で最良の値を示している．要素数問わず最良の値は太字で示している．
‡ 表記の都合で 5-tuple は 5t，metadata は mt，classtype は cl，msg は ms，reference は rf と短縮した記号を用いている．

トを作成し，機械学習モデル 5 種類を用いて適用実験を行った．SF と KF が結合された特徴量が使われることで，AAD は高精度に分類することができたが，MAD については相対的に低い精度でしか分類ができなかった．しかし，SF と WMF を結合した特徴量を使うことで，MAD において性能向上を確認した．また，有効な特徴量の分析を通して，専門家へのヒアリングから得た仮定と WMF の妥当性を確認した．

MAD におけるシグネチャを分類する際は msg と reference の 2 つが最も効果的であることから，自然言語処理分野で用いられる汎用言語モデルや単語埋め込みモデルを用いることで更なる性能向上を期待できる．

本稿の実験では，一定期間（1 年 6 カ月）内に取得したシグネチャをモデル構築・評価実験に用いた．このため，訓練時に含まれていない重要度判定に重要な単語が，判定時に未知語として現れず，テストデータの分類精度が著しく低下するという問題は起きなかった．一方，長期的に見ると，ソフトウェア情報や悪性通信の種類等の変化にともない，シグネチャに含まれる情報も変化することが想定される．この場合，重要度判定モデルの構築時に含まれていない新規単語が，判定時に未知語として表出し，判定精度を低下させる可能性がある．

シグネチャの重要度判定ミスはセキュリティ事故につながる可能性があるため，分類モデルの判定精度の低下は避けるべき事態である．このように想定した場合，分類モデルに，既定された重要度に分類するだけではない “unknown” に分類する機能があれば，より安全なセキュリティ運用に寄与することができる．その機能の実現のために，棄却オプションや [26]，能動学習での学習サンプルの選出方法（least confidence，margin sampling 等）を用いることが考えられる [27]．

長期間での重要度判定モデルの運用方法や定期的なモデ

ルの再学習方法の検討，判定時に重要な未知語が含まれる場合の影響の分析やその対処方法の検討，複数の情報通信システムでのシグネチャ分類モデルの構築・評価，情報通信システム間での分類モデルの違いの分析等は今後の課題となる．

参考文献

[1] Buczak, A.L. and Guven, E.: A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection, *IEEE Communications Surveys Tutorials*, Vol.18, No.2, pp.1153–1176 (2016).

[2] Shahriar, H. and Bond, W.: Towards an Attack Signature Generation Framework for Intrusion Detection Systems, *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress* (*DASC/PiCom/DataCom/CyberSciTech*), pp.597–603 (2017).

[3] Fallahi, N., Sami, A. and Tajbakhsh, M.: Automated flow-based rule generation for network intrusion detection systems, *2016 24th Iranian Conference on Electrical Engineering* (*ICEE*), pp.1948–1953 (2016).

[4] Lee, S., Kim, S., Lee, S., Choi, J., Yoon, H., Lee, D. and Lee, J.: LARGen: Automatic Signature Generation for Malwares Using Latent Dirichlet Allocation, *IEEE Trans. Dependable and Secure Computing*, Vol.15, No.5, pp.771–783 (2018).

[5] Constantinides, C., Shiaeles, S., Ghita, B. and Kolokotronis, N.: A Novel Online Incremental Learning Intrusion Prevention System, *2019 10th IFIP International Conference on New Technologies, Mobility and Security* (*NTMS*), pp.1–6 (2019).

[6] Chandre, P.R., Mahalle, P.N. and Shinde, G.R.: Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification, *2018 IEEE Global Conference on Wireless Computing and Networking* (*GCWCN*), pp.135–140 (2018).

[7] Tengl, S., Zhang, Z., Teng, L., Zhang, W., Zhu, H., Fang, X. and Fei, L.: A Collaborative Intrusion Detection Model using a novel optimal weight strategy based on Genetic Algorithm for Ensemble Classifier, *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design* (*CSCWD*), pp.761–766 (2018).

[8] Lotfallahtabrizi, P. and Morgan, Y.: A novel host intrusion detection system using neural network, *2018 IEEE 8th Annual Computing and Communication Workshop and Conference* (*CCWC*), pp.124–130 (2018).

[9] Moustafa, N. and Slay, J.: UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), *2015 Military Communications and Information Systems Conference* (*MilCIS*), pp.1–6 (2015).

[10] Bhuyan, M., Bhattacharyya, D.K. and Kalita, J.: Towards Generating Real-life Datasets for Network Intrusion Detection, *International Journal of Network Security*, Vol.17, pp.675–693 (2015).

[11] Chapaneri, R. and Shah, S.: A Comprehensive Survey of Machine Learning-Based Network Intrusion Detection, *Smart Intelligent Computing and Applications*, Satapathy, S.C., Bhateja, V. and Das, S. (Eds.), Singapore, pp.345–356, Springer Singapore (2019).

[12] Alsubhi, K., Al-Shaer, E. and Boutaba, R.: Alert prioritization in Intrusion Detection Systems, *NOMS 2008 - 2008 IEEE Network Operations and Management Symposium*, pp.33–40 (2008).

[13] Pietraszek, T.: Using Adaptive Alert Classification to Reduce False Positives in Intrusion Detection, *Recent Advances in Intrusion Detection*, Jonsson, E., Valdes, A. and Almgren, M. (Eds.), Berlin, Heidelberg, pp.102–124, Springer Berlin Heidelberg (2004).

[14] Stakhanova, N. and Ghorbani, A.A.: Managing Intrusion Detection Rule Sets, *Proc. 3rd European Workshop on System Security*, *EUROSEC'10*, pp.29–35 (2010).

[15] Massicotte, F. and Labiche, Y.: An analysis of signature overlaps in Intrusion Detection Systems, *2011 IEEE/IFIP 41st International Conference on Dependable Systems Networks* (*DSN*), pp.109–120 (2011).

[16] Kadhim, A.I.: Term Weighting for Feature Extraction on Twitter: A Comparison Between BM25 and TF-IDF, *2019 International Conference on Advanced Science and Engineering* (*ICOASE*), pp.124–128 (2019).

[17] Yang, Y.: Research and Realization of Internet Public Opinion Analysis Based on Improved TF - IDF Algorithm, *2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science* (*DCABES*), pp.80–83 (2017).

[18] Sun, P., Wang, L. and Xia, Q.: The Keyword Extraction of Chinese Medical Web Page Based on WF-TF-IDF Algorithm, *2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (*CyberC*), pp.193–198 (2017).

[19] Nothman, J., Qin, H. and Yurchak, R.: Stop Word Lists in Free Open-source Software Packages, *Proc. Workshop for NLP Open Source Software* (*NLP-OSS*), Melbourne, Australia, Association for Computational Linguistics (2018).

[20] Kingma, D.P. and Ba, J.: Adam: A Method for Stochastic Optimization, *3rd International Conference on Learning Representations*, *ICLR 2015*, San Diego, CA, USA, May 7–9, 2015, *Conference Track Proceedings*, Bengio, Y. and LeCun, Y. (Eds.), pp.1–15 (2015).

[21] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P.: SMOTE: Synthetic Minority over-Sampling Technique, *J. Artif. Int. Res.*, Vol.16, No.1, 321357 (2002).

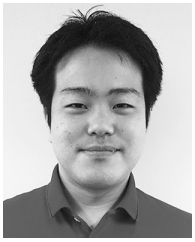
[22] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1* (*Long and Short Papers*), Minneapolis, Minnesota, pp.4171–4186, Association for Computational Linguistics (2019).

[23] Iwasaki, Y., Yamashita, A., Konno, Y. and Matsubayashi, K.: Japanese abstractive text summarization using BERT, *2019 International Conference on Technologies and Applications of Articial Intelligence* (*TAAI*), pp.1–5 (2019).

[24] Gao, Z., Feng, A., Song, X. and Wu, X.: Target-Dependent Sentiment Classification With BERT, *IEEE Access*, Vol.7, pp.154290–154299 (2019).

[25] Sohn, H. and Lee, H.: MC-BERT4HATE: Hate Speech Detection using Multi-channel BERT for Different Languages and Translations, *2019 International Conference on Data Mining Workshops* (*ICDMW*), pp.551–

559 (2019).
[26] Nadeem, M.S.A., Zucker, J.-D. and Hanczar, B.: Accuracy-Rejection Curves (ARCs) for Comparing Classification Methods with a Reject Option, *Proc. 3rd International Workshop on Machine Learning in Systems Biology*, Deroski, S., Guerts, P. and Rousu, J. (Eds.), *Proc. Machine Learning Research*, Vol.8, Ljubljana, Slovenia, PMLR, pp.65–81 (2009).
[27] Settles, B.: Active Learning Literature Survey, Computer Sciences Technical Report 1648, University of Wisconsin–Madison (2009).



川口 英俊 （正会員）

1992 年生．2014 年広島工業大学情報学部知的情報システム学科卒業．2016 年東京工業大学大学院総合理工学研究科知能システム科学専攻博士前期課程修了．同年日本電信電話株式会社に入社し在籍中．2019 年北陸先端科学技術大学院大学先端科学研究科博士後期課程に入学し在学中．機械学習の応用・実用化研究に従事．人工知能学会会員．

中谷 裕一

1980 年生．2003 年京都大学工学部情報学科卒業．2005 年京都大学大学院情報学研究科複雑系科学専攻博士前期課程修了．同年日本電信電話株式会社に入社し在籍中．2021 年筑波大学大学院システム情報工学研究群博士後期課程に入学し在学中．ネットワークセキュリティオペレーション，並列分散コンピューティングの研究に従事．電子情報通信学会会員．

岡田 将吾

1980 年生．2003 年横浜国立大学工学部卒業，2008 年東京工業大学大学院総合理工学研究科知能システム科学専攻博士後期課程修了．同年京都大学情報学研究科知能情報学専攻特定助教，2011 年東京工業大学大学院総合理工学研究科知能システム科学専攻助教．2014 年 IDIAP Research Institute 滞在研究員．2016 年東京工業大学大学院情報理工学院助教．2017 年より北陸先端科学技術大学院大学先端科学研究科准教授．博士（工学）．マルチモーダルインタラクション，人間行動解析，社会的信号処理の研究に従事．