情報処理学会論文誌 Vol.63 No.12 1699–1715 (Dec. 2022)

動的解析ログと表層情報を組み合わせたマルウェア感染活動の最終進行度推定手法

岡山 あん[1,a)] 朝倉 紗斗至[1] 中川 恒[2] 押場 博光[2] 市野 将嗣[1]

受付日 2022年3月8日, 採録日 2022年9月2日

概要：近年，マルウェアの感染を前提として，感染検知後に迅速かつ正確な対応を行うことを目指す対策が主流となりつつある．これを実現するためには，マルウェア感染後の短い挙動のログから将来的に起こりうる被害やマルウェアの挙動を推定することが必要となる．本研究ではマルウェアの将来的に起こりうる被害やそれに対応する対策をまとめたものを “最終進行度” として定義する．本論文では，回帰を用いて動的解析ログの長い挙動のログの特徴量を予測し，そして動的解析ログと表層情報を組み合わせることで，短い挙動の動的解析ログにより最終進行度を推定する手法を提案する．実験の結果，回帰を用いないで推定する場合と比較して，最大で 2.5%の精度向上に成功した．また各検体のすべてのログを学習した際に得られる精度と同程度の精度を 97.8%削減したログで実現することができた．

キーワード：マルウェア，動的解析ログ，表層情報，最終進行度推定

Final Progression Step Estimation Method of Malware Using Dynamic Analysis Logs and Surface Logs

An Okayama[1,a)] Satoshi Asakura[1] Ko Nakagawa[2] Hiromitsu Oshiba[2] Masatsugu Ichino[1]

Received: March 8, 2022, Accepted: September 2, 2022

Abstract: In recent years, on the premise of malware infection, measures aiming at prompt and accurate response after infection detection have become mainstream. In order to achieve this, it is necessary to estimate future damage and malware behavior from short logs after infection. In this research, we define “Final Progression Step” as a summary of possible future damage of malware and countermeasures against it. In this study, we use regression to predict the features of the long logs of the dynamic analysis log and the surface information and propose a method that estimates “Final Progression Step” by the dynamic analysis short logs. In the experiments, we succeeded in improving the accuracy by up to 2.5% when we compared the accuracy when we did not use regression. In addition, we were able to reduce our logs by 97.8% the number of log lines with the same accuracy which we had obtained with all the logs.

Keywords: malware, dynamic analysis logs, surface logs, final progression step estimation

1. はじめに

マルウェア対策として侵入検知システム（IDS）がよく使われている．これはネットワーク上を流れるパケットをすべて監視する機能を有するものや，定期的にログを取得，監視を行い，管理者へ警告を送るものがあげられる [1]．これを利用し，多くの場合は通知された不正なアクセス情報やパケット情報をもとに通信を遮断する．また近年は攻撃の巧妙化が進み，マルウェアの侵入を完全に防ぐことが難しくなっている．一般にインシデントレスポンスは，事前準備，検知・分析，封じ込め・根絶・復旧，事後対応の四段階に分けられる [2]．攻撃者からの攻撃をたとえ 1 度早期に検知し遮断できたとしても，攻撃者はその後さらに高度

[1] 電気通信大学
The University of Electro-Communications, Chofu, Tokyo 182–0021, Japan
[2] 株式会社 FFRI セキュリティ
FFRI Security, Inc., Chiyoda, Tokyo 100–0005, Japan
[a)] a.okayama@uec.ac.jp

な攻撃を仕掛けてくる可能性が考えられるため，感染を検知し遮断できたとしても，次の攻撃に対する事前準備が必要となってくる．ただしIDSがマルウェアによる感染を検知するまでの時間は攻撃によって異なるため，それに応じて得られる挙動のログの長さも異なる．つまり，早期に検知できた場合，インシデントレスポンスの事前準備段階で使えるログが少量しか存在しないことが考えられる．よって生成される挙動ログの行数が様々であったとしても，攻撃者の意図を少しでも理解するための事前準備ができるようにする必要がある．それをふまえ，本研究では，将来，マルウェアの感染により起こりうる挙動や被害，その対策をまとめたものを“最終進行度”として定義して，これを短い挙動の動的解析ログから推定することを目指す．最終進行度推定はIDSで検知し遮断するまでのログを利用することを想定しているため，一定の長さの動的解析ログを得るまでマルウェアを泳がすのではなく，遮断するまでに得られた様々な長さのログからそれぞれ推定する．また感染後初期に行う挙動が似ているマルウェアが存在するため，得られた短い挙動のログでは，その後起こりうる挙動の判断が難しい場合が考えられる．そこでその場合でも最終進行度を推定できるよう，短い挙動のログからその後得られるはずだったログを予測する工夫を行った．なお推定した最終進行度は，マルウェア対策を考える際の優先順位を決める際や，対策が必要な範囲をしぼる際に利用できると考える．

本論文では，動的解析ログと表層情報を組み合わせた最終進行度推定を提案する．動的解析ログはマルウェアの挙動を把握するのに適しているため，最終進行度推定に利用した．また表層情報は，端末に被害が生じる前に取得することができ，環境に依存せずに取得することが可能という利点がある．そのため表層情報を利用した．そして，マルウェアの動的解析ログと表層情報に基づき，定義した最終進行度にマルウェア分類することで，最終進行度を推定した．

なお本論文は，マルウェア対策研究人材育成ワークショップ2021で発表した論文[3]をさらに発展させたものである．

以下2章で関連研究，3章で予備実験，4章で提案手法，5章で実験方法，6章で実験結果，7章で考察，8章でまとめと今後の課題を示す．

2. 関連研究

本論文では，マルウェアの短い動的解析ログと表層情報を用いた最終進行度推定手法を提案する．本研究では，マルウェアにより起こりうる被害やそれに対して必要な対策をまとめたものを最終進行度として定義した．そのため本章では，マルウェアのリスク分析，リスク予測に関する既存研究，マルウェアの早期推定を行っている既存研究，進行度に関連する既存研究について述べ，最後に本研究の位置づけを説明する．

2.1 リスク分析，リスク予測に関する研究

Leylaら[4]は端末の過去1年間のバイナリファイルを解析することで，リスクレベルを定量化し，どの端末が感染の危険にさらされているのかを予測した．実験の結果，TPRが96%，FPRが5%の結果を得た．Kichangら[5]は悪意のあるAPIデータベースと比較することで，Androidアプリケーションのセキュリティリスクを評価する方法を提案した．実験の結果，90%以上の精度で良性，悪性アプリを分類することができた．西野ら[6]は攻撃者のネットワークへの通信試行ログを用いて，そのログから攻撃が進み深刻な被害が生じる高リスク時間帯なのか，軽微な被害にとどまる低リスク時間なのかを分類した．実験の結果，93%の精度を得た．矢野ら[7]は西野らの研究の“DNNをベースにしているために，モデルの解釈性が乏しい”という問題点を受けて，作成したDNNモデルを最も良く近似できる線形分類器を作成することで，問題の解決を図った．実験の結果，97%の精度を得た．

2.2 マルウェアの早期推定を行っている研究

Yuvrajら[8]は静的解析を行って悪意のあるバイナリを抽出し，その後ATT&CK[9]のフレームワークにマッピングすることで，リアルタイムで多段階の攻撃を検知する方法を提案した．実験の結果，98%の精度でマルウェアを検知することができた．Niteshら[10]は4秒の動的解析ログと，静的解析ログそれぞれにおいてマルウェアのクラス分類を行った．実験の結果，静的解析ログを用いた場合には97%，動的解析ログを用いた場合には99%の精度を得た．Matildaら[11]は検体を実行してから4秒のログを用いて，その検体が悪意のあるものか否かを予測した．実験の結果，91%の精度を得た．朝倉ら[12]は記録時間の短い動的解析ログから得られる特徴量を用いて，記録時間の長い動的解析ログから得られる特徴量を予測することで，マルウェアの機能推定を行った．実験の結果，記録開始から5秒までのログに対し，特徴量の予測を行わない場合よりも2.6%精度向上することを示した．

2.3 進行度に関連する研究

Sudhirら[13]はマルウェアの感染が進行する段階を12段階に分け，各ステージにおけるマルウェアの挙動を監視することで，マルウェアか否かを分類した．実験の結果，97%の精度を得た．寺田ら[14]はマルウェアの活動を，ネットワーク通信の観点で8つのフェーズに整理して，それぞれのフェーズの遷移を整理した遷移モデルを定義した．その後，マルウェア通信データを各フェーズにあてはめ，最終的にマルウェアの種別を推定した．その結果，公開データを用いた場合はほぼGereric Trojanに誤分類した

が，BOS Dataset [15] を用いた場合はすべての検体に対して正しく分類された．

2.4 本研究の位置づけ

本研究は，できるだけ短い行数の動的解析ログで最終進行度推定を可能とすることを研究目的としている．

2.1 節で述べた文献 [4], [5] は，マルウェアのリスク分析であり，短い動的解析ログでの推定を想定した手法ではなかった．文献 [6], [7] は用いるデータがネットワーク通信ログである点と，研究の焦点が短い動的解析ログでの推定ではない点が本研究とは異なる．また 2.2 節で述べた文献 [8], [10], [11], [12] は，研究目的がそれぞれマルウェア検知，マルウェアのクラス分類，マルウェアの機能推定であり，本研究の最終進行度推定とは異なる．2.3 節の文献 [13], [14] は，焦点が短い動的解析ログでの推定ではない点，研究目的がマルウェアの検知，もしくはマルウェアのクラス分類である点が本研究とは異なる．

3. 予備実験

最終進行度推定において使用する動的解析ログの行数と分類精度の関係を評価することで，より長い行数のログを使用するほど推定精度が高くなるということを確認する．なお本研究は，早期に検知し遮断された場合を使用シーンとして想定しているため，“行数” は動的解析ログの 1 行目から数えた行数を指す．

また表層情報が最終進行度ごとに異なった特徴を持つことと，実際に分類精度を確認することで，最終進行度推定において表層情報が有用であることを確認する．

3.1 使用データ

実験には MWS Datasets の一部として提供される Soliton Dataset 2020 [15] を使用した．これは株式会社ソリトンシステムズのエンタープライズ向け EDR 製品である InfoTrace Mark II for Cyber（以降 “Mark II” と呼ぶ）が導入された環境にて取得されたログのデータセットである．Cuckoo Sandbox 上に Windows 7 Pro ベースで Mark II を導入したゲスト端末でマルウェアを実行して記録した Mark II ログ，Cuckoo ログが提供されている．Soliton Dataset 2020 には，検体それぞれに対して表層情報も提供されている．本研究の分類実験では，同データセットで提供されている全 581 検体の Mark II ログと表層情報のうち，3.2 節で説明するラベル付け方法で最終進行度ラベルが付与できた 554 検体の Mark II ログと表層情報を使用した．Soliton Dataset 2020 に付属されている情報をもとに，実験で使用した 554 検体に含まれるマルウェアファミリ，もしくは関連するマルウェアの名称を**表 1** にまとめる．

表 1 マルウェアファミリ名・関連するマルウェア名
Table 1 Malware family name·related malware name.

ファミリ名・関連するマルウェア名	検体数	ファミリ名・関連するマルウェア名	検体数
RoyalRoad	106	Emotet	66
Evasive Azorult	25	NetWire	25
Predator the Thief	25	STOP Ransomware	25
Valyria	25	マイナードロッパー	25
Upatre	24	XtremeRAT	18
HOPLIGHT	16	TSCookie	16
Novter	14	FTCODE	12
Nanocore	12	DRIDEX	11
Ramnit	11	njRAT	10
Bifrost	8	Daper	8
Tick(BronzerButler)	8	PurpleFox	7
DarkHotel	6	MegaCortex	6
AZORULT	5	RETADUP	4
Ryuk	4	BlackTech(HUAPI)	3
Ryuk Stealer	3	FAREIT	2
IceDown	2	GenericKD	2
Trickbot	2	PochC2	2
Nemty	2	Advanced IP Scanner	1
AGENTTESLA	1	Ammyy Admin	1
ARTFULPIE	1	BISTROMATH	1
BUFFETLINE	1	Coinminer.Win.64.MALXMR.TIAOODBZ	1
CROWDEDFLOUNDER	1	EKANS	1
HackTool.Win32.Impacket.AI	1	HOTCROISSANT	1
KPortScan3	1	LockerGoga	1
Mailto	1	mRemoteNC	1
mRemoteNG	1	MS16-032	1
NLBrute1.2	1	NS.exe	1
Parallax RAT	1	Pierogi	1
RancorDLL	1	RancorVBScript	1
SLICKSHOES	1	SoftPerfect Network Scanner	1
Spark	1	Trickbot ローダー	1
Trojan.PS1.LUDICROUZ.A	1	Trojan.PS1.PCASTLE.B	1
Trojan.Win32.CVE20178464.A	1	Trojan.Win32.DLOADR.AUSUPY	1
TrojanSpy.Win32.BEAHNY.THCACAI	1	Wiper	1
Worm.Win32.BLASQUI.A	1	xDedicLogCleaner	1
unknown	1		

3.2 最終進行度ラベルについて

Soliton Dataset 2020 で利用された Cuckoo2.0.7 [16] には，ATT&CK の各シグネチャに対して TTPID が付与されており，同データセットでこの情報が記録されているので，これを用いて最終進行度ラベルへのラベル付けを行った．まず ATT&CK で使われている用語について 3.2.1 項で説明し，3.2.2 項で詳細な最終進行度ラベルの付与方法を説明する．

3.2.1 ATT&CK

ATT&CK とは，MITRE 社が開発している攻撃者の攻撃手法および体系をまとめたフレームワーク・ナレッジベースである．これは “Adversary Group（攻撃者）”，“Software（攻撃ツール）”，“Techniques（技術）”，“Tactics（戦術）” の 4 つの観点でまとめられている．また ATT&CK の公式ホームページには，各 Tactics に対して用いられる具体的な Techniques がまとめられている Matrix と呼ばれる表がある．なお各 Tactics，Techniques には ID が付与されており，これを TTPID と呼ぶ．

3.2.2 最終進行度ラベルの付与方法

最終進行度ラベルを付与するために，まず各検体の Cuckoo ログから付与された TTPID をすべて抽出し，それらと ATT&CK の Matrix を照らし合わせ，各 TTPID を対応する Tactics に変換した．次に，ATT&CK の Matrix の情報を参考にマルウェアに感染したあとに起こりうるマルウェアの挙動や被害の観点から進行度を独自に定義し

表 2 進行度の定義
Table 2 Definition of the progression.

進行度	該当する Tactics
1	Initial Access, Execution, Persistence, Privilege Escalation
2	Defense Evasion
3	Credential Access, Discovery
4	Lateral Movement, Collection
5	Command and Control, Exfiltration, Impact

表 3 最終進行度の内容
Table 3 Detail of the final progression.

最終進行度	該当する説明
1	初期感染動作で停止する可能性あり 今後攻撃が進行する可能性があるため，端末の動作ログと通信ログに異常があるか監視する必要あり
2	検知を避ける挙動をしたのち，停止する可能性あり 今後攻撃が進行する可能性があるため，端末の動作ログと通信ログに異常があるか監視する必要あり
3	アカウント情報や，環境情報が盗まれる可能性あり ID,PW を見直したり，端末内，周辺機器の構成の環境を悪用されないように見直す必要あり
4	ファイルの構成情報の奪取や，感染の拡大を行おうとする可能性あり 重要ファイルの置き場所の再検討や，環境から感染端末を切り離す必要あり
5	C&C サーバとの通信や，重要ファイルの奪取を行おうとする可能性あり 重要ファイルの置き場所の再検討や，環境から感染端末を切り離す必要あり

表 4 ラベルごとの検体数
Table 4 Number of data per label.

ラベル	検体数
最終進行度 1	5
最終進行度 2	264
最終進行度 3	249
最終進行度 4	31
最終進行度 5	5

た．これを**表 2** に示す．そして各検体で出現した Tactics と表 2 を照らし合わせ，各 Tactics を進行度に変換した．その後，変換した進行度のうち最大の進行度を最終進行度ラベルとした．各最終進行度ラベルの説明を**表 3** に記載した．そして，実験で使用した 554 検体に対して，最終進行度ラベルの付与を行い，各最終進行度ラベルに対する検体数は**表 4** に示すとおりとなった．

3.2.3 単語抽出，特徴ベクトルの作成方法

（1）に Mark II ログを使用した際の単語抽出，特徴ベクトルの作成，（2）に表層情報を使用した際の単語抽出，特徴ベクトルの作成方法を述べる．なお，Mark II ログ，表層情報を組み合わせた分類を行う際には．それぞれで作成した特徴ベクトルを列方向に結合した特徴ベクトルを使用した．

（1）Mark II ログを用いた場合

本研究では，Mark II ログからイベント（ログ中の evt），サブイベント（ログ中の subEvt）を抽出し，コロンでつなげたものを単語として使用した．使用した単語は**表 5** にまとめた 27 種類である．なお得られる挙動の情報量を増やすために，イベントが “file”，サブイベントが “close” である場合は，ファイルの読み込みのバイト数が書き込みのバイト数より多い場合は “file:read”，少ない場合は “file:write” とした．また等しい場合は，“file:close” とした．なおこれは，朝倉らの単語抽出手法 [12] を参考に行った．この結果，Mark II ログから 27 種類の単語を得た．次に全検体に出現する単語を網羅するコーパスを作成した．最後に各検体の Mark II ログの evt，subEvt の 1 組を 1 行と定義し，はじめから n 行目までの Mark II ログ（以降 “n 行ログ” と呼ぶ）に登場する単語からコーパスに記載のある単語の出現頻度を算出した．

表 5 使用した単語
Table 5 Words used.

使用した単語			
file:del	file:delDir	file:close	file:read
file:write	file:create	file:rename	file:make
file:download	file:chgAttr	file:copy	win:active
net:dcon	net:con	net:webURL	net:openUDP
ps:start	ps:stop	ps:inject	reg:create
reg:setVal	reg:delVal	reg:delKey	session:loginR
sys:startRed	sys:start	sys:stop	

本研究では，n 行ログからコーパスに記載のある単語数分の要素を持つベクトルを作成し，それを分類に用いる特徴量として用いた．なお n 行ログを使用する際に m $(< n)$ 行しかない検体に関しては，m 行目までを利用して単語抽出した．この際も名称としては n 行ログとしている．

（2）表層情報を用いた場合

表層情報から特徴量を抽出する際には，株式会社 FFRI セキュリティが提供している表層情報抽出ツールである FEXRD [17] を利用した．その結果，抽出した 2,347 種類の表層情報を単語とした．これを網羅するコーパスを作成したのち，各単語に対して抽出した値を要素とした特徴ベクトルを作成し，これを分類に用いる特徴量とした．

3.2.4 実験方法

3.2.3 項のとおりに，n 行ログと表層情報のそれぞれから作成した特徴ベクトルを列方向に結合した．学習とテストには，これを特徴ベクトルとして使用した．まず scikit-learn [18] に含まれる StandardScaler を用いて特徴量を標準化し，得られた特徴量を学習データとテストデータとして用いて RandomForestClassifier（以降 “RF” と呼ぶ）で分類を行った．なお分類は，層化 5 分割交差検証を行った．この際，各交差検証で各最終進行度ラベルの検体数の比率をそろえるために StratifiedKFold を用いた．

分類に用いる特徴量については選択して利用した．RF のパラメータを**表 6** の範囲で，GridSearch を利用して適切なパラメータ値を探した．そして全単語 2,374 種類（Mark II ログ：27，表層情報：2,347）のうち，SelectFromModel で重要度が中央値以上である単語のうち 46 個選択し，その後，選択した 46 個の特徴量を要素とした特徴ベクトルに対して再度 GridSearch を用いて適切なパラメータ値を

表 6 実験で用いたパラメータ 1
Table 6 Parameter with experiment 1.

	パラメータ	範囲
RF	n_estimators	100
	max_depth	6,7,8,9,10
	criterion	gini, entropy

表 7 Mark II ログのみを用いた予備実験
Table 7 Preliminary experiment with Mark II logs.

使用した行数	1	10	20	30	40	50	...	全行数使用
分類精度（%）	49.8	57.0	57.9	61.4	85.6	87.4	...	91.9

表 8 Mark II ログ，表層情報を用いた予備実験
Table 8 Preliminary experiment with Mark II Logs and surface logs.

使用した行数	1	10	20	30	40	50	...	全行数使用
分類精度（%）	84.1	84.7	84.8	87.9	89.0	90.4	...	92.1

探した．なお本研究では，各交差検証時の学習データに対してパラメータチューニングを行う際の指標として，StratifiedKFold を用いた層化 3 分割交差検証で RF を用いて分類を行った際の accuracy を用いた．そして全検体に対して，分類結果である予測ラベルを最終進行度として出力した．

3.3 Mark II ログに対する予備実験

最終進行度推定において使用する Mark II ログの行数と分類精度の関係を評価することで，より長い行数のログを使用するほど推定精度が高くなるということを確認する．またその際，3.2.4 項で示した実験方法をもとに，Mark II ログのみを使用した分類実験と，表層情報と組み合わせた分類実験を行った．

3.2.3 項より各検体の Mark II ログの evt，subEvt の 1 単語を 1 行，1 行目から n 行目までの Mark II ログを n 行ログと定義した．学習，テストに用いた Mark II ログの n 行ログの n を “使用した行数”，それを使用して実験を行った際に得られた正答率を “分類精度” として，Mark II ログのみを用いた分類結果を**表 7**，表層情報と組み合わせた分類結果を**表 8** にまとめた．

2 つの分類結果より，使用した行数が増えるほど，分類精度が高くなるという傾向が確認できた．

3.4 表層情報に対する予備実験

表層情報が最終進行度ごとに異なった特徴を持つことと，表層情報のみを用いた場合の分類精度を確認することで，最終進行度推定において表層情報が有用であることを確認する．

3.4.1 表層情報の特徴

表層情報を分析したところ，3.2.2 項で定義する 5 段階の進行度で表現される最終進行度ラベルごとに差異が見られた．例として，各最終進行度ラベルの file_size の中央値を**表 9** に示す．この表より，最終進行度ラベルごとに差異が見られることが分かる．file_size 以外の表層情報にも各最終進行度ラベルで異なる特徴が見られた．

表 9 file_size の特徴
Table 9 Features of file_size.

最終進行度ラベル	中央値
1	28797.0
2	465143.5
3	454656.0
4	262144.0
5	2310858.0

3.4.2 表層情報の分類精度

3.2.4 項で示した分類実験を，動的解析ログを用いず表層情報のみを用いて行った．その結果，83.8%の分類精度を得た．

4. 提案手法

3 章で示した予備実験では，使用する動的解析ログの行数を長くすると分類精度が高くなること，表層情報が最終進行度ごとに異なった特徴を持つこと，そして実際に表層情報のみを用いた分類実験の結果から表層情報が最終進行度推定において有用であることが分かった．ここで本研究の目的は，できるだけ短い行数の動的解析ログで高精度に最終進行度推定を行うことである．そのために（i），（ii）の 2 点工夫した．

なお提案手法では，2 つの異なる行数のログを比較した際に短い行数のログを “短いログ”，長い行数のログを “長いログ” と表記する．たとえば，20 行ログから作成した特徴量を 80 行ログから作成した特徴量に予測する場合には，20 行のログが短いログ，80 行のログが長いログとなる．また 10 行ログから作成した特徴量を 20 行ログから作成した特徴量に予測する場合には，10 行のログが短いログ，20 行のログが長いログとなる．このように 20 行のログが，比較対象によって短いログ，長いログのどちらにもなりうる．

（i）短いログから回帰モデルを用いて長い行数の動的解析ログの特徴量を予測する手法を利用

朝倉ら [12] は，マルウェアの機能推定において，短いログから長いログでの特徴量を予測した手法を提案し，有効性を示した．さらに，3 章で示した予備実験で，より長い行数のログを使用するほど推定精度が高くなるということを確認した．これより，短いログの特徴量から長いログの特徴量が予測できれば短いログで高精度に最終進行度推定を行うことが可能となると考えられる．以上より，1 点目の工夫として，短いログから得られる特徴量を説明変数とし，長いログから得られる特徴量を目的変数とした回帰モデルで，長いログから得られる特徴量を予測し，その後それを用いて分類を行った．

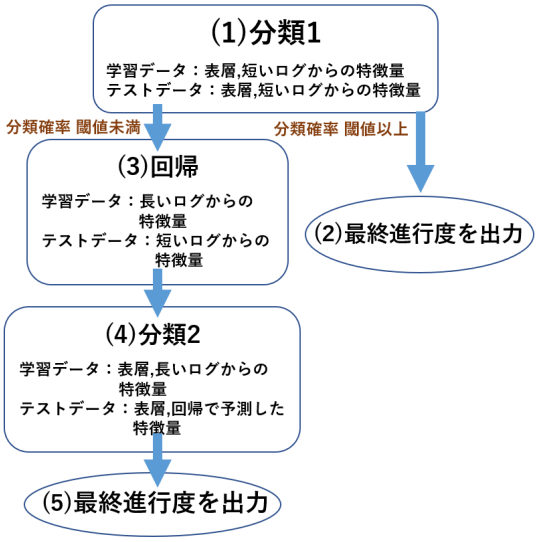


図 1 提案手法
Fig. 1 Proposed method.

(ii) 表層情報の利用

実験に用いた動的解析ログは，初期段階で多くの検体がプロセスの起動を行っており，用いるログの行数を増やすほどレジストリやファイルへの操作の記録が増えるため，ログの行数が長くなるほど最終進行度ごとに差が現れるという特徴があった．3 章で示したが，予備実験として，分類に用いるログの行数を増やした場合の分類精度の変化を評価したところ，ログの行数を増やすほど分類精度が向上するという結果が得られた．しかし本研究では，できるだけ短いログを用いることに焦点を当てている．そのため最終進行度ごとで差異が少ない短いログを用いた場合でも，高精度に分類する必要がある．そこで本論文では表層情報を利用することにした．表層情報は 3.4 節で示したとおり，定義する 5 段階の進行度で表現される最終進行度ラベルごとに差異がみられた．また実際に 3.3 節では，Mark II ログと表層情報を組み合わせたことで精度が向上したことを確認した．そのため，表層情報が最終進行度推定に有用であると考え，利用した．

上記の (i)，(ii) の工夫点を盛り込んだ最終進行度推定手法を提案する．具体的には以下に示す (1) から (5) の手順で構成され，図示すると**図 1** である．なお本論文では，分類器が出力する各最終進行度に対する所属確率を “分類確率” と定義し．そのうち最も分類確率が高い最終進行度を “予測ラベル” とする．

(1) すべての検体の短いログから得られる特徴ベクトルと表層情報から得られる特徴ベクトルを組み合わせた特徴量に対して，学習，テストデータ用いて分類を行う．なおこの 1 回目の分類を，以降は “分類 1” と呼ぶ．
(2) 分類 1 で判定した予測ラベルの分類確率が閾値以上の検体のみ，その予測ラベルを最終進行度とする．
(3) 分類 1 で判定した予測ラベルの分類確率が閾値未満の検体に対しては，長いログから得られる特徴量を用いて学習した回帰モデルで短いログから得られる特徴量を予測する．
(4) 分類 1 で判定した予測ラベルの分類確率が閾値未満の検体に対して，長いログから得られる特徴ベクトルと表層情報から得られる特徴ベクトルを組み合わせた特徴量を学習データ，短いログから得られる特徴ベクトルから回帰で長いログから得られる特徴ベクトルに予測した特徴ベクトルと，表層情報から得られる特徴ベクトルを組み合わせた特徴量をテストデータとして分類を行う．なおこの 2 回目の分類を，以降は “分類 2” と呼ぶ．
(5) 分類 2 を行った検体に対して，分類 2 で判定した予測ラベルを最終進行度とする．

分類 1 で判定した予測ラベルの分類確率が閾値以上の検体に関しては，分類 1 の時点で判定した予測ラベルへの信頼度が高いと判断し，判定した予測ラベルを最終進行度とする．一方で閾値未満の検体に関しては，正しい分類を行うのに信頼度が高くないと判断し，分類 2 を実施し，分類 2 で判定した予測ラベルを最終進行度とする．

なお，マルウェア実行後の初期の短い行数のログを用いて最終進行度推定を行うにあたって，最大の進行度を示した挙動が実験に使用する初期のログに含まれる場合と，含まれない場合の 2 つが起こりうる．含まれる場合は，実験によって “現時点よりもマルウェアの感染が進むか否か” が分かり，初期のログに含まれない場合は，“現時点より将来的にどのくらいマルウェアの感染が進むか” が分かる．そのためどちらの場合であっても，最終進行度を予測する意義がある．

5. 実験

本実験では，提案手法を用いれば，(本来，高い分類精度を実現するには多いログが必要であったが) 少ない動的ログの行数であっても高い分類精度を実現できることを示す．なお実験方法全体を**図 2** に示す．

5.1 実験方法

本実験で使用したアルゴリズムなどを，図 2 に記載した順に説明する．なお，図の同じ色の特徴ベクトルは，同一の特徴ベクトルであることを表す．

また，使用したパラメータの探索範囲は表 6，**表 10** のとおりである．

5.1.1 分類 1

3.2.4 項で示した分類を，全検体に対して行った．また分類時には RF に用意されている predict_proba を用いて予測ラベルの分類確率を算出し，それが閾値以上であった検体については，分類結果である予測ラベルを最終進行度として出力した．この際の閾値としては，55，60，65%を

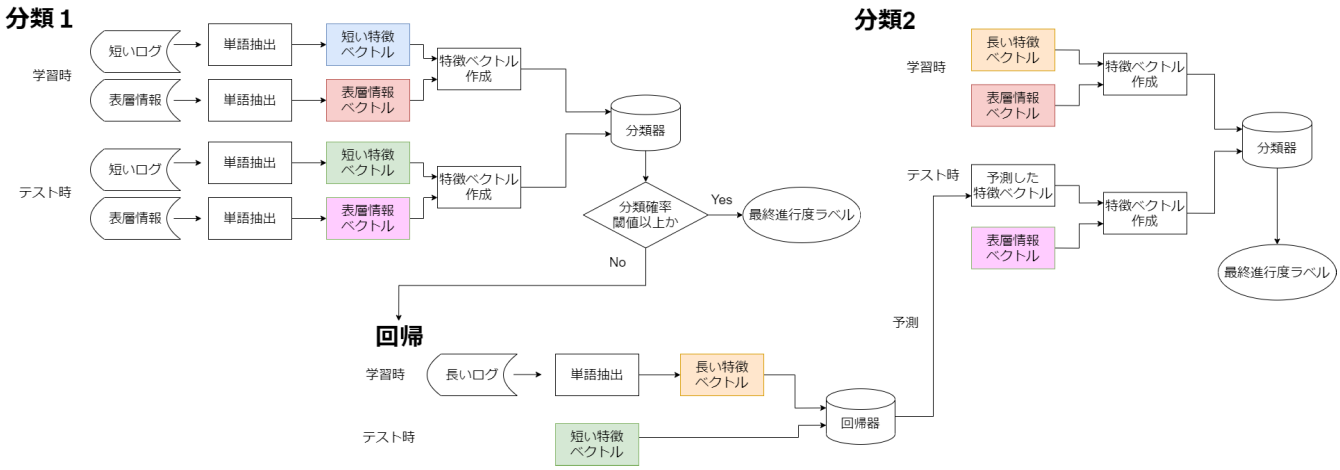


図 2 実験方法
Fig. 2 Experiment method.

表 10 実験で用いたパラメータ 2
Table 10 Parameter with experiment 2.

	パラメータ	範囲
RFR	n_estimators	16, 17, 18, 19, 20
	max_depth	13, 14, 15, 16, 17, None
DT	max_depth	7, 8, 9, 10, 11
	criterion	gini, entropy

使用した．

5.1.2 回帰

5.1.1 項で，予測ラベルの分類確率が閾値未満であった検体については，Mark II ログから得られる特徴量を回帰で予測した．以下，詳細な方法を説明する．まず StandardScaler を用いて特徴量を標準化したのちに RandomForestRegressor（以降 “RFR” と呼ぶ）を用いて，StratifiedKFold の層化 5 分割交差検証で長いログから得られる特徴量を予測した．

本研究の回帰の目的は，回帰後に行う分類の精度向上であるため，各交差検証時の学習データに対してパラメータチューニングを行う際の指標として，3.2.4 項の分類を RF の代わりに DecisionTreeClassifier（以降 “DT” と呼ぶ）を用いて StratifiedKFold の層化 4 分割交差検証で行った際の accuracy を用いた．

回帰モデルの学習時には，長いログから得られる特徴量を目的変数とし，短いログから得られる特徴量を説明変数とした．5.1.3 項で説明する分類 2 には，学習した回帰モデルで，短いログから得られる特徴量を用いて長いログから得られる特徴量を予測したものを用いた．

5.1.3 分類 2

回帰モデルを用いて特徴量を予測した後，再度分類を行った．なおこの際の学習データは長いログから得られる特徴ベクトルと表層情報から得られる特徴ベクトルを組み合わせた特徴量，テストデータは短いログから得られる特徴ベクトルを用いて長いログから得られる特徴ベクトルに予測したものと，表層情報から得られる特徴ベクトルを組み合わせた特徴量である．また使用した分類方法，パラメータなどは 3.2.4 項と同じものである．

5.2 予備実験との差異

予備実験は，3.2.4 項で示した方法で全検体に対して分類実験を行い，予測結果を最終進行度とした．そのため回帰，分類 2 は行っていない．なお予備実験の結果は表 7，表 8 に示したとおりである．それに対して，本実験では，まず全検体に対して予備実験と同様の分類実験（分類 1）を行った．そしてその際の分類確率が閾値以上であった検体については予測結果を最終進行度とし，閾値未満であった検体については回帰を適用した後，さらに分類実験（分類 2）を行い，その予測結果を最終進行度とした．つまり，本実験は一部検体に回帰を適用し，分類 2 を行ったという点が，予備実験との差異である．よって予備実験と本実験の精度を比較することで，一部検体に回帰を適用し，分類 2 を行ったことの有効性を確認することができる．

5.3 評価方法

提案手法の有効性を以下の 2 つの観点で評価した．

(1) 短いログを使用したときの精度向上
(2) 必要なログの削減

(1) に関しては，短いログを使用したときの精度向上を目的に，回帰の説明変数に用いる短いログから得られる特徴量を固定として，回帰の目的変数に用いるログの行数を増やしていく実験を行う．例として短いログを 15 行，長いログを 20 行使用する場合を考える．15 行ログを使用して全検体に対して分類 1 のみで最終進行度推定を行った精度を，一部検体に対して 15 行ログから得られる特徴量を用いて予測した 20 行ログから得られる特徴量を使用した提案手法の精度が上回ることができれば，一部検体に回帰を適用することによって精度向上が見込めたといえる．

表 11 Mark II ログのみを用いた最終進行度推定 1（短いログ固定）
Table 11 Final progression estimation 1 with Mark II logs (Fixed short log).

閾値 55%	使用した行数	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/275	—	166	166	166	166	166	166	166	166	166
	分類 2 の正解数/279	—	154	107	122	139	149	158	156	150	153
	分類精度（%）	57.9	57.8	49.2	52.0	55.1	56.9	**58.5**	**58.1**	57.0	57.6
	使用した行数		120	130	140	150	160	170	180	190	200
	分類 1 の正解数/275		166	166	166	166	166	166	166	166	166
	分類 2 の正解数/279		156	156	156	155	156	156	156	156	156
	分類精度（%）		**58.1**	**58.1**	**58.1**	**58.1**	57.9	**58.1**	**58.1**	**58.1**	**58.1**
閾値 60%	使用した行数	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/76	—	62	62	62	62	62	62	62	62	62
	分類 2 の正解数/478	—	252	210	208	190	187	220	250	239	251
	分類精度（%）	57.9	56.7	49.1	48.7	45.5	44.9	50.9	56.3	54.3	56.5
	使用した行数		120	130	140	150	160	170	180	190	200
	分類 1 の正解数/76		62	62	62	62	62	62	62	62	62
	分類 2 の正解数/478		251	257	254	256	252	255	254	255	255
	分類精度（%）		56.5	57.6	57.0	57.4	56.7	57.2	57.0	57.2	57.2
閾値 65%	使用した行数	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/75	—	61	61	61	61	61	61	61	61	61
	分類 2 の正解数/479	—	259	207	207	202	189	222	259	237	256
	分類精度（%）	57.9	57.8	48.4	48.4	47.5	45.1	51.1	57.8	54.0	57.2
	使用した行数		120	130	140	150	160	170	180	190	200
	分類 1 の正解数/75		61	61	61	61	61	61	61	61	61
	分類 2 の正解数/479		259	255	259	256	256	257	260	256	255
	分類精度（%）		57.8	57.0	57.8	57.2	57.2	57.4	57.9	57.2	57.2

(2) に関しては，できるだけ短いログを用いた最終進行度推定を意識して，必要なログの削減を目的に行う．そのため，実験としては，回帰の目的変数に用いる長いログから得られる特徴量を固定として，回帰の説明変数に用いるログの行数を減らしていくことを行った．例として短いログを 30 行，長いログを 1,000 行使用する場合を考える．このとき 1,000 行ログを使用して全検体に対して分類 1 のみで最終進行度推定を行った精度と，提案手法の 30 行ログから得られる特徴量を 1,000 行から得られる特徴量に予測したものを用いた精度が等しければ，1,000 行必要な精度を 30 行で得ることができたといえる．このときのログの削減率は，$(1000 - 30)/1000 = 0.97$ と計算する．

6. 実験結果

6.1 Mark II ログのみを用いた実験

Mark II ログと表層情報を組み合わせる有効性を示すために，5.1 節の提案手法の実験を Mark II ログのみで行った．

5.3 節 (1)，(2) のとおり，提案手法の評価には 2 種類の観点で実験を行った．まず精度向上をみる (1) では，閾値を 55，60，65%，短いログとして 10，15，20 行のログで実験をした際，最も精度が良かったケースは閾値を 55%，短いログとして 20 行を使用した結果であった．ここで短いログを 20 行と固定して閾値を変えた際の結果を**表 11**，閾値を 55%に固定して短いログを変えた際の結果を**表 12** に示す．なお表では，短いログ（表 11 では 20 行，表 12 では 10，15，20 行）から予測する予測対象のログの行数を“使用した行数”，それを使用して実験を行った際に得られた正答率を “分類精度”，分類を正解であった検体数を “正解数” とした．各表の一番左にある短いログ（10，15，20 行のログ）の分類精度は分類のみで精度評価を行った予備実験の精度であるため，全 554 検体に関して分類を行い，精度評価を行っている．そのため提案手法の分類 1，2 とは正解数の分母が異なるため，分類 1，2 の正解数を横線とした．またこの精度を超えると提案手法によって回帰を用いることで精度向上がみられたといえる．該当する精度を太字にしている．表 11，表 12 をみると，太字の精度のうち閾値を 55%，短いログを 20 行としたとき，長いログが 80 行の 58.5%のときが最も高いことが分かる．つまりこれは，提案手法によってベースとなる 20 行の精度 57.9%から，最大 0.6%精度向上がみられたことがいえる．

またログの削減割合をみる 5.3 節で示した (2) では，閾値を 55，60，65%，長いログとして各検体すべてのログを使用したもので実験をした．このときの結果を**表 13** に示す．なお表では，全行数に予測するのに用いた短いログの行数を “使用した行数”，それを使用して実験を行った際に得られた正答率を “分類精度”，分類を正解であった検体数を “正解数” とした．また 5.3 節で示した (1) とは異なり，分類 1 で用いるログが固定ではないため，表 13 の分類 1，2 の正解数の分母である “各分類を行った検体数” は異なる．ここで全行数を使用した精度 91.9%と同程度以上の精度を探すため，母比率の区間推定を行う．母比率とは，各要素の特性が特定の状態 A であるか否かのいずれかのとき状態 A である要素の割合を指す [19]．大きさ m の標本か

表 12 Mark II ログのみを用いた最終進行度推定 1（閾値固定）
Table 12 Final progression estimation 1 with Mark II logs (Fixed threshold).

	使用した行数	10	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/211	—	131	131	131	131	131	131	131	131	131	131
	分類 2 の正解数/343	—	180	180	137	151	136	137	165	180	171	181
	分類精度（%）	57.0	56.1	56.1	48.4	50.9	48.2	48.4	53.4	56.1	54.5	56.3
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/211		131	131	131	131	131	131	131	131	131	
	分類 2 の正解数/343		182	184	184	183	184	184	184	184	182	
	分類精度（%）		56.5	56.9	56.9	56.7	56.9	56.9	56.9	56.9	56.5	
	使用した行数	15	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/262	—	162	162	162	162	162	162	162	162	162	162
	分類 2 の正解数/292	—	157	157	133	131	126	113	140	149	148	152
	分類精度（%）	57.9	57.6	57.6	53.2	52.9	52.0	49.6	54.5	56.1	56.0	56.7
	使用した行数		120	130	140	150	160	170	180	190	200	
閾値 55%	分類 1 の正解数/262		162	162	162	162	162	162	162	162	162	
	分類 2 の正解数/292		144	157	157	157	157	157	153	157	157	
	分類精度（%）		55.2	57.6	57.6	57.6	57.6	57.6	56.9	57.6	57.6	
	使用した行数	20		30	40	50	60	70	80	90	100	110
	分類 1 の正解数/275	—		166	166	166	166	166	166	166	166	166
	分類 2 の正解数/279	—		154	107	122	139	149	158	156	150	153
	分類精度（%）	57.9		57.8	49.2	52.0	55.1	56.9	**58.5**	**58.1**	57.0	57.6
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/275		166	166	166	166	166	166	166	166	166	
	分類 2 の正解数/279		156	156	156	155	156	156	156	156	156	
	分類精度（%）		**58.1**	**58.1**	**58.1**	**58.1**	57.9	**58.1**	**58.1**	**58.1**	**58.1**	

表 13 Mark II ログのみを用いた最終進行度推定 2（長いログ固定）
Table 13 Final progression estimation 2 with Mark II logs (Fixed long log).

	使用した行数	10	20	30	40	50	60	70	80	90	100	
	分類 1 の正解数	131	166	308	463	468	473	490	501	492	490	
	分類 2 の正解数	179	136	33	9	14	16	11	4	6	8	
	分類精度（%）	56.0	54.5	61.6	85.2	87.0	88.3	**90.4**	**91.2**	**89.9**	**89.9**	
閾値 55%	使用した行数	200	300	400	500	600	700	800	900	1000	...	全行数使用
	分類 1 の正解数	488	490	487	488	486	486	488	494	494		—
	分類 2 の正解数	11	14	13	13	11	14	12	10	8		—
	分類精度（%）	**90.1**	**91.0**	**90.3**	**90.4**	**89.7**	**90.3**	**90.3**	**91.0**	**90.6**	...	91.9
	使用した行数	10	20	30	40	50	60	70	80	90	100	
	分類 1 の正解数	61	62	118	440	457	459	484	487	484	481	
	分類 2 の正解数	241	232	211	33	19	20	18	10	11	19	
	分類精度（%）	54.5	53.1	59.4	85.4	85.9	86.5	**90.6**	**89.7**	89.4	**90.3**	
閾値 60%	使用した行数	200	300	400	500	600	700	800	900	1000	...	全行数使用
	分類 1 の正解数	487	480	482	482	482	482	482	484	485		—
	分類 2 の正解数	14	15	13	19	15	13	19	17	15		—
	分類精度（%）	**90.4**	89.4	89.4	**90.4**	**89.7**	**89.4**	**90.4**	**90.4**	**90.3**	...	91.9
	使用した行数	10	20	30	40	50	60	70	80	90	100	
	分類 1 の正解数	61	61	90	418	446	450	471	476	473	475	
	分類 2 の正解数	231	239	218	54	23	23	25	22	22	27	
	分類精度（%）	52.7	54.2	55.6	85.2	84.7	85.4	89.5	**89.9**	89.4	**90.6**	
閾値 65%	使用した行数	200	300	400	500	600	700	800	900	1000	...	全行数使用
	分類 1 の正解数	472	476	471	475	473	474	472	479	471		—
	分類 2 の正解数	27	25	23	22	14	21	22	21	24		—
	分類精度（%）	**90.1**	**90.4**	89.2	**89.7**	**87.9**	89.4	89.2	**90.3**	89.4	...	91.9

ら得られた標本比率の実現値を $\overline{x}$ とすると，母比率 p の信頼区間は，

$$\overline{x} - t\sqrt{\frac{\overline{x}(1-\overline{x})}{m}} \leqq p \leqq \overline{x} + t\sqrt{\frac{\overline{x}(1-\overline{x})}{m}}$$

と示される [19]．なお t は，確率変数 Z が標準正規分布 $N(0,1)$ に従う際に $P(Z > t)$ を満たす値とする．今回は検体が最終進行度推定を正しくできたか否を状態 A とする．554 検体の標本から得られた標本比率の実現値 $\overline{x}$ は，91.9%より 0.919 である．また信頼率 95%で信頼区間を推定するとき，t は 1.96 となる．今回は 91.9%と同程度以上の精度を探したいため，95%の信頼区間の下限を計算すると，

表 14 Mark II ログ，表層情報を用いた最終進行度推定 1（短いログ固定）
Table 14 Final progression estimation 1 with Mark II logs and surface logs (Fixed short log).

閾値 55%	使用した行数	10	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/509	—	446	446	446	446	446	446	446	446	446	446
	分類 2 の正解数/45	—	22	22	20	23	17	20	23	22	17	18
	分類精度（%）	84.7	84.5	84.5	84.1	84.7	83.6	84.1	84.7	84.5	83.6	83.8
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/509		446	446	446	446	446	446	446	446	446	
	分類 2 の正解数/45		20	17	18	17	17	18	17	17	17	
	分類精度（%）		84.1	83.6	83.8	83.6	83.6	83.8	83.6	83.6	83.6	
閾値 60%	使用した行数	10	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/485	—	436	436	436	436	436	436	436	436	436	436
	分類 2 の正解数/69	—	40	37	38	38	40	40	41	38	40	39
	分類精度（%）	84.7	**85.9**	**85.4**	**85.6**	**85.6**	**85.9**	**85.9**	**86.1**	**85.6**	**85.9**	**85.7**
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/485		436	436	436	436	436	436	436	436	436	
	分類 2 の正解数/69		42	41	42	41	39	40	39	39	38	
	分類精度（%）		**86.3**	**86.1**	**86.3**	**86.1**	**85.7**	**85.9**	**85.7**	**85.7**	**85.6**	
閾値 65%	使用した行数	10	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/453	—	415	415	415	415	415	415	415	415	415	415
	分類 2 の正解数/101	—	64	62	64	64	62	64	68	65	67	65
	分類精度（%）	84.7	**86.5**	**86.1**	**86.5**	**86.5**	**86.1**	**86.5**	**87.2**	**86.6**	**87.0**	**86.6**
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/453		415	415	415	415	415	415	415	415	415	
	分類 2 の正解数/101		66	65	65	65	66	66	66	66	63	
	分類精度（%）		**86.8**	**86.6**	**86.6**	**86.6**	**86.8**	**86.8**	**86.8**	**86.8**	**86.3**	

$$0.919 - 1.96\sqrt{\frac{0.919(1-0.919)}{554}} \fallingdotseq 0.896$$

となった．よって 89.6%以上を同程度と見なし，表 13 で該当する精度を太字にした．このうち最も短いログの行数は閾値 55%，60%のときの 70 行と分かる．

6.2 提案手法の実験

同様に 5.1 節の実験を，Mark II ログと表層情報を用いて，5.3 節 (1)，(2) の 2 種類の観点で実験を行った．

精度向上を見る (1) では，閾値を 55，60，65%，短いログとして 10，15，20 行のログで実験をして，分類精度を評価した．その結果，最も精度が良かったケースである閾値が 65%，短いログとして 10 行を使用した結果であった．ここで閾値を 65%に固定して短いログを変えた際の結果を**表 14**，短いログを 10 行と固定して閾値を変えた際の結果を**表 15** に示す．なお表では，短いログ（表 14 では 10 行，表 15 では 10，15，20 行）から予測する予測対象のログの行数を “使用した行数”，それを使用して実験を行った際に得られた正答率を “分類精度”，分類を正解であった検体数を “正解数” とした．各表の一番左にある短いログ（10，15，20 行のログ）の分類精度は分類のみで精度評価を行った際の予備実験の精度であるため，全 554 検体に関して分類を行い，精度評価を行っている．そのため提案手法の分類 1，2 とは正解数の分母が異なるため，分類 1，2 の正解数を横線とした．また 6.1 節と同様，ベースとなる短いログの分類精度を超えている行は精度向上がみられるため，太字で示している．表 14，表 15 をみると，太文字の精度のうち，閾値を 65%，短いログを 10 行としたとき，長いログが 80 行の 87.2%が最も精度が高いといえる．よって最大で 2.5%の精度が向上したといえる．

次に 5.3 節で示した (2) でログの削減割合を確認する．閾値を 55，60，65%，長いログとして各検体すべてのログを使用したもので実験をした．このときの結果を**表 16** に示す．なお表では，全行数に予測するのに用いた短いログの行数を “使用した行数”，それを使用して実験を行った際に得られた正答率を “分類精度”，分類を正解であった検体数を “正解数” とした．また 5.3 節で示した (1) とは異なり，分類 1 で用いるログが固定ではないため，表 16 の分類 1，2 の正解数の分母である “各分類を行った検体数” は異なる．6.1 節と同様に，全行数を使用した精度 92.1%と同程度以上の精度を探すため，母比率の 95%信頼区間の下限は，

$$0.921 - 1.96\sqrt{\frac{0.921(1-0.921)}{554}} \fallingdotseq 0.900$$

となった．よって 90.0%以上の精度を出したものを各検体の全行数を使用した場合の精度と同程度と見なし，表 16 で該当する精度を太字で表した．このうち最も短い行数は閾値 55%，60%のときの 60 行であった．

今回は提案手法に対して Mark II ログを用いた実験を行い，Mark II ログの場合，有効性を確認することができた．

表 15 Mark II ログ，表層情報を用いた最終進行度推定 1（閾値固定）
Table 15 Final progression estimation 1 with Mark II logs and surface logs (Fixed threshold).

	使用した行数	10	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/453	—	415	415	415	415	415	415	415	415	415	415
	分類 2 の正解数/101	—	64	62	64	64	62	64	68	65	67	65
	分類精度（%）	84.7	**86.5**	**86.1**	**86.5**	**86.5**	**86.1**	**86.5**	**87.2**	**86.6**	**87.0**	**86.6**
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/453		415	415	415	415	415	415	415	415	415	
	分類 2 の正解数/101		66	65	65	65	66	66	66	66	63	
	分類精度（%）		**86.8**	**86.6**	**86.6**	**86.6**	**86.8**	**86.8**	**86.8**	**86.8**	**86.3**	
	使用した行数	15	20	30	40	50	60	70	80	90	100	110
	分類 1 の正解数/455	—	418	418	418	418	418	418	418	418	418	418
	分類 2 の正解数/99	—	59	58	59	57	57	57	53	57	56	59
	分類精度（%）	84.7	**86.1**	**85.9**	**86.1**	**85.7**	**85.7**	**85.7**	**85.0**	**85.7**	**85.6**	**86.1**
	使用した行数		120	130	140	150	160	170	180	190	200	
閾値 65%	分類 1 の正解数/455		418	418	418	418	418	418	418	418	418	
	分類 2 の正解数/99		57	57	58	60	57	56	59	57	55	
	分類精度（%）		**85.7**	**85.7**	**85.9**	**86.3**	**85.7**	**85.6**	**86.1**	**85.7**	**85.4**	
	使用した行数	20		30	40	50	60	70	80	90	100	110
	分類 1 の正解数/446	—		402	402	402	402	402	402	402	402	402
	分類 2 の正解数/108	—		66	63	62	61	62	60	58	58	58
	分類精度（%）	84.8		84.5	83.9	83.8	83.6	83.8	83.4	83.0	83.0	83.0
	使用した行数		120	130	140	150	160	170	180	190	200	
	分類 1 の正解数/446		402	402	402	402	402	402	402	402	402	
	分類 2 の正解数/108		57	57	57	60	58	59	61	62	61	
	分類精度（%）		82.9	82.9	82.9	83.4	83.0	83.2	83.6	83.8	83.6	

表 16 Mark II ログ，表層情報を用いた最終進行度推定 2（長いログ固定）
Table 16 Final progression estimation 2 with Mark II logs and surface logs (Fixed long log).

	使用した行数	10	20	30	40	50	60	70	80	90	100	
	分類 1 の正解数	446	436	462	469	476	486	497	495	493	491	
	分類 2 の正解数	20	29	19	23	19	14	15	11	13	13	
	分類精度（%）	84.1	83.9	86.8	88.8	89.4	**90.3**	**92.4**	**91.3**	**91.3**	**91.0**	
閾値 55%	使用した行数	200	300	400	500	600	700	800	900	1000	...	全行数使用
	分類 1 の正解数	493	497	496	499	497	496	496	500	499		—
	分類 2 の正解数	16	10	13	10	14	12	15	9	10		—
	分類精度（%）	**91.9**	**91.5**	**91.9**	**91.9**	**92.2**	**91.7**	**92.2**	**91.9**	**91.9**	...	92.1
	使用した行数	10	20	30	40	50	60	70	80	90	100	
	分類 1 の正解数	436	416	445	458	462	479	487	487	487	482	
	分類 2 の正解数	35	38	36	29	29	26	22	25	26	28	
	分類精度（%）	85.0	81.9	86.8	87.9	88.6	**91.2**	**91.9**	**92.4**	**92.6**	**92.1**	
閾値 60%	使用した行数	200	300	400	500	600	700	800	900	1000	...	全行数使用
	分類 1 の正解数	485	491	491	491	487	488	491	495	493		—
	分類 2 の正解数	19	18	18	14	18	20	14	13	18		—
	分類精度（%）	**91.0**	**91.9**	**91.9**	**91.2**	**91.2**	**91.7**	**91.2**	**91.7**	**92.2**	...	92.1
	使用した行数	10	20	30	40	50	60	70	80	90	100	
	分類 1 の正解数	415	402	434	443	447	466	477	478	475	472	
	分類 2 の正解数	67	56	52	30	42	31	28	29	23	30	
	分類精度（%）	87.0	82.7	87.7	85.4	88.3	89.7	**91.2**	**91.5**	89.9	**90.6**	
閾値 65%	使用した行数	200	300	400	500	600	700	800	900	1000	...	全行数使用
	分類 1 の正解数	474	478	482	480	472	477	478	482	482		—
	分類 2 の正解数	29	26	18	22	30	31	28	33	24		—
	分類精度（%）	**90.8**	**91.0**	**90.3**	**90.6**	**90.6**	**91.7**	**91.3**	**93.0**	**91.3**	...	92.1

7. 考察

本章では，最終進行度推定に表層情報を用いたことによる有効性，短いログの特徴量から長いログの特徴量を予測したことによる有効性の 2 つの観点から実験結果を考察する．

7.1 表層情報を用いたことによる有効性

7.1.1 分類精度向上への貢献度

表 11，表 12，表 14，表 15 をみると，Mark II ログのみ

表 17 分類時の feature_importance
Table 17 Feature importance of classification.

	使用した行数	10	20	30	40	50	60	70	80	90	100
Mark II ログ	種類数/27	2	2	6	7	9	10	10	10	11	10
	feature_importance 合計	0.36	0.21	0.80	0.78	1.0	1.5	1.4	1.6	1.6	1.5
	feature_importance 平均	**0.18**	**0.10**	**0.13**	**0.11**	**0.11**	**0.15**	**0.14**	**0.16**	**0.15**	**0.15**
表層情報	種類数/2347	44	44	40	39	37	36	36	36	35	36
	feature_importance 合計	4.0	4.2	3.9	4.0	3.6	3.2	3.1	3.0	2.8	3.0
	feature_importance 平均	**0.092**	**0.095**	**0.098**	**0.10**	**0.098**	**0.089**	**0.087**	**0.083**	**0.081**	**0.082**
	使用した行数	110	120	130	140	150	160	170	180	190	200
Mark II ログ	種類数/27	12	11	12	12	12	13	12	13	12	13
	feature_importance 合計	1.7	1.7	1.6	1.6	1.6	1.7	1.7	1.8	1.7	1.7
	feature_importance 平均	**0.14**	**0.15**	**0.13**	**0.14**	**0.13**	**0.13**	**0.14**	**0.14**	**0.15**	**0.13**
表層情報	種類数/2,347	34	35	34	34	34	33	34	33	34	33
	feature_importance 合計	2.9	2.8	2.8	2.8	2.9	2.7	2.8	2.8	2.7	2.7
	feature_importance 平均	**0.084**	**0.081**	**0.083**	**0.084**	**0.085**	**0.083**	**0.083**	**0.083**	**0.080**	**0.083**

を用いた場合では全体的に 57〜58%という分類精度であったのに対し，表層情報と組み合わせることで全体的に 86〜87%に分類精度を大きく向上させた．また 3.4.2 項より，表層情報のみを用いた実験では 83.8%の分類精度を得た．よって Mark II ログと表層情報を組み合わせることで，最大 3.4%の精度向上を得たといえる．

また Mark II ログと表層情報を組み合わせた分類時に，どの単語が重要視されているのかを RF の feature_importance [18] を用いて出力した．なおこの分類実験は，5.1.1 項で示した分類 1 と同様に行った．分類に用いた行数と，各分類時で分類モデルに採用された Mark II ログから得られる単語と表層情報から得られる単語の種類数，各ログの feature_importance の合計値（Mark II ログの全単語 27 種類の合計値，表層情報の全単語 2,347 種類の合計値），そしてその値を種類数で割った平均値を，**表 17** にまとめる．なお表では，実験に使用した n 行ログの n を “使用した行数” とした．まず Mark II ログから得られる単語と表層情報から得られる単語のうち，使用した行数 10 行から 200 行までの場合において登場した Mark II ログと表層情報の単語の頻度の割合を計算した．Mark II ログから得られる単語の登場回数の合計値は 199 であり，表層情報から得られる単語の登場回数の合計値は 721 であった．よって Mark II ログ，表層情報それぞれから得られる単語の頻度の割合は 0.22，0.78 となった．また各単語の feature_importance の合計値の割合も計算した．Mark II ログから得られる単語の feature_importance の合計値は 35.96，表層情報から得られる単語の feature_importance の合計値は 64.04 であった．よって Mark II ログ，表層情報それぞれから得られる単語の feature_importance の合計値の割合は 0.36，0.64 となる．以上より，単語の頻度の割合，単語の feature_importance の合計値のいずれの観点からも表層情報が高く重要視されていると判断できる．

さらに Mark II ログと表層情報では，作成した単語の総種類数に差異が見られるため，feature_importance の合計値を登場した単語の種類数で割った平均値での比較も行った．これを行うことで，Mark II ログの feature_importance の合計値を 27（種類），表層情報の feature_importance の合計値を 2,347（種類）で割った場合よりも，実験に使用したデータセットに即した単語の種類数の違いに考慮して，feature_importance を確認することができる．この値は表 17 より，Mark II ログは 0.10〜0.18，表層情報は 0.080〜0.10 の値をとることが分かる．これより，表層情報を Mark II ログと遜色なく重要視していることが分かる．

以上の点から，表層情報と Mark II ログを組み合わせることは，分類精度向上に十分に寄与していることが分かる．

7.1.2 各予測ラベルの精度向上への貢献度

本提案では，短いログを用いた分類精度が低いという問題を表層情報と組み合わせることで改善を試みた．表層情報と組み合わせることで得られた各予測ラベルの精度向上への貢献度の変化を考察するため，Mark II ログから得られるすべての特徴量のみを用いた分類，Mark II ログから得られるすべての特徴量と表層情報から得られるすべての特徴量を組み合わせた分類，Mark II ログから得られるすべての特徴量と表層情報の特徴量を 1 つ組み合わせた分類の 3 種類の分類実験を行った．これにより，Mark II ログと表層情報を組み合わせることが，各予測ラベルの分類精度にどのように影響を及ぼし，また表層情報がどの程度有効であったかを確認する．なおこの分類実験は，5.1.1 項で示した分類 1 と同様の方法で行った．ここで 3 種類の分類実験において，それぞれの分類時の各予測ラベルに対する精度を**表 18** に示す．なお表の “精度” は，各最終進行度ラベルの正解数を表 4 の検体数で割った正解率を表している．また表では，Mark II ログから得られるすべての特徴量に組み合わせる 1 つの表層情報の特徴量の例として，“file_size” の値，“strings_char_hist[45]” [*1]の値，

[*1] FEXRD を用いて抽出された strings_char_hist 配列の 45 番目の値．

表 18 Mark II と表層情報を用いた場合の精度
Table 18 Accuracy with Mark II logs and surface logs.

		最終進行度 1	最終進行度 2	最終進行度 3	最終進行度 4	最終進行度 5
Mark II ログのみ	10 行	0%	0.95%	0.27%	0%	0%
	20 行	0%	0.95%	0.28%	0%	0%
Mark II ログ+すべての表層情報	10 行	0.2%	0.91%	0.80%	0.87%	0%
	20 行	0.2%	0.89%	0.83%	0.84%	0%
Mark II ログ+file_size	10 行	0%	0.84%	0.88%	0.58%	0%
	20 行	0%	0.84%	0.77%	0.55%	0%
Mark II ログ+strings_char_hist[45]	10 行	0%	0.89%	0.72%	0.35%	0.2%
	20 行	0%	0.89%	0.71%	0.45%	0.2%
Mark II ログ+imports_hint[82]	10 行	0%	0.93%	0.31%	0.84%	0.40%
	20 行	0%	0.93%	0.30%	0.81%	0%

“imports_hint[82]”[*2]の値を用いた場合を示している．これらの特徴量は，Mark II ログから得られるすべての特徴量と表層情報から得られるすべての特徴量を用いた分類実験の際に，feature_importance が高かった表層情報の特徴量である．

表 18 をみると，まず Mark II ログのみを用いた場合では，主に最終進行度 2 と分類されており，他の最終進行度ラベルの精度が低いことが分かる．またすべての表層情報と組み合わせることで，最終進行度 2 だけでなく，最終進行度 1，3，4 の分類精度が向上していることが分かる．さらに Mark II ログから得られる特徴量に，分類時に高く重要視されていた file_size の値を組み合わせるだけで，最終進行度 3，4 の精度が向上することが分かった．同様に高く重要視されていた strings_char_hist [45] の値と Mark II ログから得られる特徴量，imports_hint [82] の値と Mark II ログから得られる特徴量を組み合わせたそれぞれの実験では，ともに最終進行度 3，4，5 の精度向上がみられる．

以上の点から，表層情報と Mark II ログを組み合わせることは，各予測ラベルの精度向上に寄与することが分かる．

7.1.3 ログの削減への貢献度

ログの削減に表層情報がどの程度寄与したか考察する．なお本実験で使用した Soliton Dataset 2020 の 554 検体の総ログ行数の平均は，2,679.5 行である．

6.1 節より，Mark II ログのみを用いた実験では，89.6%以上の分類精度を満たす最も短い行数は閾値 55%，60%のときの 70 行と分かった．つまりログの全行数を使用する場合と 70 行のみを使用する場合では同程度の分類精度が得られているといえるので，ログの削減割合は，$(2679.5 - 70)/2679.5 \fallingdotseq 97.4\%$ と計算できる．また 6.2 節より，表層情報と Mark II ログを組み合わせた実験では，90.0%以上の分類精度を出した最も短い行数は，閾値 55%，60%のときの 60 行と分かった．このときの削減割合は，$(2679.5 - 60)/2679.5 \fallingdotseq 97.8\%$ と分かる．

以上の点から，表層情報を取り入れたことによってさらに 0.4%削減でき，表層情報と Mark II ログを組み合わせることは，ログ削減に寄与することが分かった．

7.2 短いログの特徴量から長いログの特徴量を予測したことによる有効性

7.2.1 分類精度向上，ログ削減への貢献度

回帰を用いて短いログの特徴量から長いログの特徴量を予測したことが，どの程度分類精度の向上に寄与したか考察する．表 14，表 15 において，最大の精度を得た 10 行のログを回帰で 70 行までの特徴量を閾値 65%で予測した場合を例として考える．この際，2.5%の精度向上に成功した．結果について分析すると，分類 1 で不正解，分類 2 で正解であった検体は 33 検体であり，逆に分類 1 で正解，分類 2 で不正解であった検体は 21 検体であった．この差である 12 検体分が，分類 1 の結果を予測ラベルとせず分類 2 の結果を予測ラベルとしたことによって，分類結果が改善されたマルウェアに該当する．一方で提案手法では，最終進行度 1，5 の多くの検体を正しく分類することができなかった．Mark II ログを見ると最終進行度 1，5 と他の最終進行度との差異が生じるには 40 行程度が必要である．つまり，回帰を用いて 70 行のときの特徴量を予測したこの実験では本来正しく分類できるはずである．しかしながら，実際は回帰による予測精度が不十分であるため，正しく分類することができなかったと考える．

次に回帰を用いて短いログの特徴量から長いログの特徴量を予測したことが，ログの削減にどの程度寄与したかを考察する．6.2 節では，60 行使用したときの分類精度が全行数を使ったときの分類精度と同程度であったため，97.8%ログを削減することに成功したと述べた．Mark II ログの平均行数は 2,679.5 行であり，これを 60 行で同程度の精度で推定できたのは意義があったと考える．

以上の点から，短いログの特徴量から長いログの特徴量を予測したことは，精度向上に寄与し，またログの削減に大きく寄与したといえる．

7.2.2 予測性能の評価

回帰器による特徴量の予測部分の評価を行った．表層情

[*2] FEXRD を用いて抽出された imports_hint 配列の 82 番目の値．

表 19 20 行ログから作成した特徴量を予測する予測対象のログの行数と RMSE の関係（閾値 55%）
Table 19 Relationship between the number of target lines predicted from the feature of 20lines and RMSE (Threshold 55%).

予測対象の行数	30	40	50	60	70	80	90	100	110
RMSE	0.15	0.71	1.3	1.8	2.3	2.8	3.1	3.5	3.9
予測対象の行数	120	130	140	150	160	170	180	190	200
RMSE	4.3	4.6	5.0	5.4	5.8	6.2	6.6	7.0	7.5

報の影響を無視するため，Mark II ログのみを用いた実験結果である表 11，表 12 と，表 13 に対応した実験の予測性能を評価した．評価方法は，実際に得られる特徴量の値と予測により得られる特徴量の値との誤差を二乗平均平方根誤差（以降，“RMSE”）を計算して確認した．なお RMSE は，scikit-learn で用意されている metrics.mean_squared_error で計算した平均二乗誤差に平方根をつけて計算したものである．また，RMSE の値のみでは予測が正しく行うことができたのかという判断が難しいため，さらに，長いログから作成した特徴量のみを用いた分類（予備実験）で正解した検体を対象に，短いログから作成した特徴量を長いログから作成した特徴量に予測した特徴量を用いた分類（提案手法）で正解しているかを確認することで，予測の正しさを確認した．特徴量の予測を正しく行うことができれば，短いログから作成した特徴量を長いログから作成した特徴量に予測した特徴量を用いた分類（提案手法）で正解する検体数が，長いログから作成した特徴量のみを用いた分類（予備実験）で正解した検体数と同程度になると考えたためである．

まず表 11，表 12 に注目する．ここで最大の精度 58.5%を得た閾値 55%において，20 行ログの特徴量から予測する対象の行数と予測した特徴量の RMSE の関係を**表 19** に示す．この表をみると，行数が増えるほど RMSE が増加していることが分かる．これは予測する行数が増えるほど，予測した特徴量の値が大きくなるためである．

また，最大の精度 58.5%を得た 20 行ログから作成した特徴量を回帰で 80 行ログから作成した特徴量に予測した場合（分類 1 の閾値 55%）を例として考える．予測が正しくできていれば，20 行ログから作成した特徴量を 80 行ログから作成した特徴量に予測した特徴量を用いた分類と，80 行ログから作成した特徴量のみを用いた分類は精度が同程度であると考えられる．回帰，分類 2 を行った 279 検体に対して，20 行ログから作成した特徴量を 80 行から作成した特徴量に予測した際の分類結果と，回帰を用いず 80 行ログから作成した特徴量のみを使用した際の分類結果の関係を**表 20** にまとめる．これより 144 検体は正しく分類することができ，105 検体は正しく分類できなかったことが分かる．つまり，80 行ログで正しく分類できた（144 + 105 =）249 検体中 144 検体については，正しく分類できる程度の

表 20 20 行ログから作成した特徴量を 80 行から作成した特徴量に，予測した場合と 80 行ログから作成した特徴量を使用した場合の分類結果
Table 20 Classification result when predicted the features to ones of 80lines from the ones of 20lines logs and using the ones of 80lines logs.

		80 行ログの分類結果	
		正	誤
20 行ログの特徴量から	正	144	14
80 行に予測した分類結果	誤	105	16

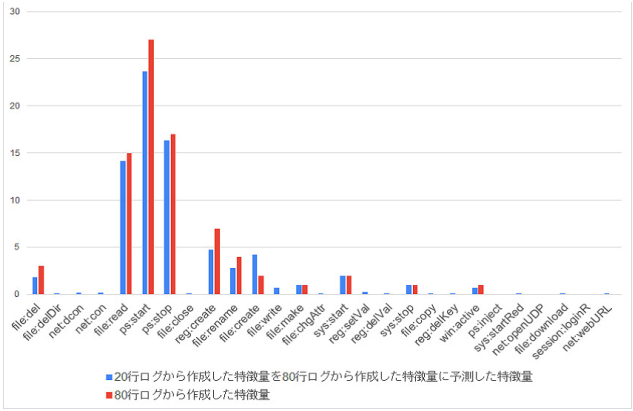


図 3 80 行ログから作成した特徴量を使用した分類で正解，20 行ログから作成した特徴量を 80 行から作成した特徴量に予測した際の分類で正解であった検体
Fig. 3 Correct answer in using the features of 80lines, correct answer when predicted the features to ones of 80lines from ones of 20lines.

予測ができていると考える．ここで，80 行ログから作成した特徴量を使用した分類で正解，20 行ログから作成した特徴量を 80 行ログから作成した特徴量に予測した際の分類でも正解であった 1 検体と，80 行ログから作成した特徴量を使用した分類では正解，20 行ログから作成した特徴量を 80 行から作成した特徴量に予測した際の分類で不正解であった 1 検体の特徴量の予測結果を**図 3**，**図 4** に表す．横軸が特徴量，縦軸が各特徴量の値であり，青が 20 行ログから作成した特徴量を 80 行から作成した特徴量に予測した特徴量，赤が 80 行から作成した特徴量を表す．図 3 をみると，80 行ログから作成した特徴量を使用した分類で正解，20 行ログから作成した特徴量を 80 行ログから作成した特徴量に予測した際の分類でも正解であった検体ということもあり，80 行ログから作成した特徴量の値と予測した特徴量の値の多くが同程度であり，正しく予測できていることが分かる．また図 4 をみると，予測した特徴量の値と実際の 80 行ログから作成した特徴量の値が似ていないことが分かる．この検体は赤い数値と同じ値に予測できていれば正解できた検体である．つまり正しく特徴量が予測できていないために，不正解であったと考えられる．

次に表 13 で，最大の精度 91.2%を得た閾値 55%におい

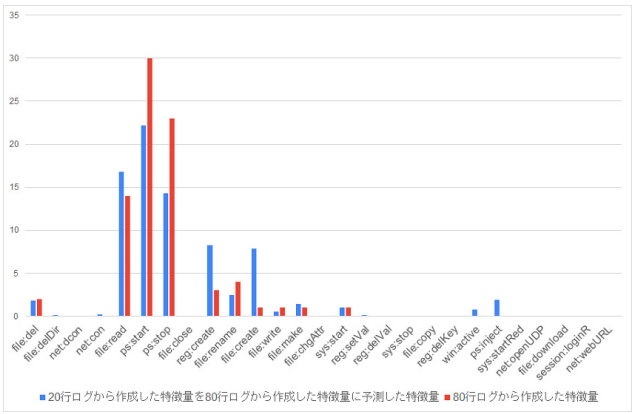


図 4 80 行ログから作成した特徴量を使用した分類で正解，20 行ログから作成した特徴量を 80 行から作成した特徴量に予測した際の分類で不正解であった検体

Fig. 4 Correct answer in using the features of 80lines, incorrect answer when predicted the features to ones of 80lines from ones of 20lines.

表 21 全行数のログから作成した特徴量に予測する短いログの行数と RMSE の関係（閾値 55%）

Table 21 Relationship between the number of short lines to predict for the features of all lines and RMSE (Threshold 55%).

短いログの行数	10	20	30	40	50	60	70	80	90	100
RMSE	1438	1521	2087	2474	2775	1922	1763	534	649	1819
短いログの行数	200	300	400	500	600	700	800	900	1000	
RMSE	1948	3322	3021	2987	2892	2577	3110	1597	3280	

て，全行数のログから作成した特徴量に予測する短いログの行数と予測した特徴量の RMSE の関係を**表 21** に示す．この表をみると，表 19 とは異なり，短いログの行数が増えるほど RMSE が増加するとはいえないことが分かる．表 21 は予測する対象の行数が全行数と固定されているため，使用する短いログの行数によらず予測した特徴量の値の差異が大きくないことが要因であると考える．

また最大の精度 91.2%を得た 80 行ログから作成した特徴量を回帰で全行数のログから作成した特徴量に予測した場合（分類 1 の閾値 55%）を例として考える．予測を正しくできていれば，80 行ログから作成した特徴量を全行数のログを作成した特徴量に予測した特徴量を用いた分類と，全行数のログから作成した特徴量のみを用いた分類は精度が同程度であると考えられる．回帰，分類 2 を行った 11 検体に対して，80 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した際の分類結果と，回帰を用いず全行数のログから作成した特徴量のみを使用した際の分類結果の関係を**表 22** にまとめる．これより正しく予測できた検体はなかったことが分かる．表 13 を確認すると，他の行数の場合と比べて多くの検体が分類 1 で分類確率が閾値を超え，回帰，分類 2 に回らなかったことが分かる．回帰，分類 2 に回った検体は予測，分類が難しい検体が多かったために，正しく予測できなかったと考える．こ

表 22 80 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した場合と全行数のログから作成した特徴量を使用した分類結果

Table 22 Classification result when predicted the features to ones of all lines from ones of 80lines logs and using the features of all lines logs.

		全行数のログの分類結果	
		正	誤
80 行ログの特徴量から全行数予測した分類結果	正	0	4
	誤	3	4

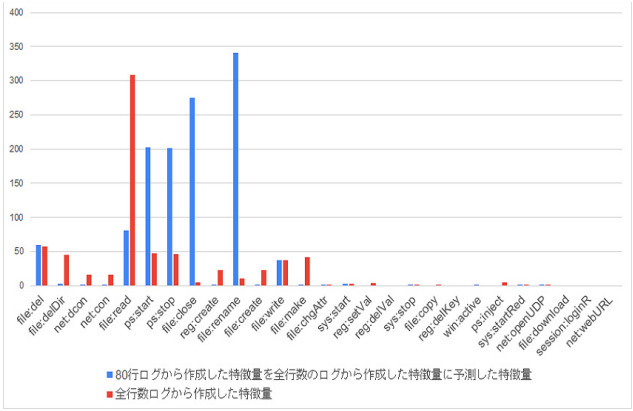


図 5 全行数のログから作成した特徴量を使用した分類で正解，80 行ログから作成した特徴量を全行数から作成した特徴量に予測した際の分類で不正解であった検体

Fig. 5 Correct answer in using the features of all lines, inCorrect answer when predicted the features to ones of all lines from ones of 80lines.

こで全行数のログから作成した特徴量のみを使用した分類で不正解，80 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した特徴量を使用した分類で不正解であった検体のうち，1 検体の予測結果を**図 5** に表す．横軸が特徴量，縦軸が各特徴量の値であり，青が 80 行ログの特徴量を全行数のログから作成した特徴量に予測した特徴量，赤が全行数のログから作成した特徴量を表す．図 5 をみると，赤と青の値の差異が大きくなっていることが分かる．正しく特徴量が予測できていないために，不正解であったと考えられる．

また表 13 において，次に高い精度である 91.0%を得た 300 行ログから作成した特徴量を回帰で全行数のログから作成した特徴量に予測した場合（分類 1 の閾値 55%）を例として考える．回帰，分類 2 を行った 31 検体に対して，300 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した際の分類結果と，回帰を用いず全行数のログから作成した特徴量のみを使用した際の分類結果の関係を**表 23** にまとめる．これより 8 検体は正しく分類でき，6 検体は正しく分類できなかったことが分かる．つまり，全行数のログで正しく分類できた（8＋6＝）14 検体中 8 検体については，正しく分類できる程度の予測ができてい

表 23 300 行ログから作成した特徴量を全行数から作成した特徴量に予測した場合と全行数のログから作成した特徴量を使用した場合の分類結果

Table 23 Classification result when predicted the features to one of all lines from one of 300lines logs and using the features of all lines logs.

		全行数のログの分類結果	
		正	誤
300 行ログの特徴量から全行数予測した分類結果	正	8	6
	誤	6	11

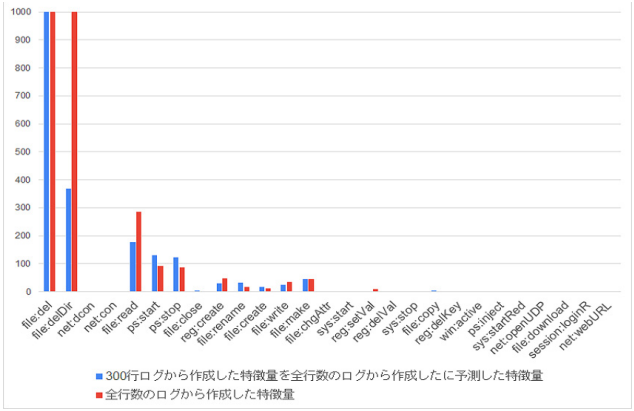


図 6 全行数のログから作成した特徴量を使用した分類で正解，300 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した際の分類で正解した検体

Fig. 6 Correct answer in using the features of all lines, correct answer when predicted the features to ones of all lines from ones of 300lines.

ると考えられる．ここで，全行数のログから作成した特徴量を使用した分類で正解，300 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した際の分類でも正解であった 1 検体と，全行数のログから作成した特徴量を使用した分類では正解，300 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した際の分類で不正解であった 1 検体の特徴量の予測結果を**図 6**，**図 7** に表す．横軸が特徴量，縦軸が各特徴量の値であり，青が 300 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した特徴量，赤が全行数のログから作成した特徴量を表す．なお図 6 の file:del（赤，青），file:delDir（赤），図 7 の file:del（赤）の縦軸の値は 1,000 以上であるが，他の特徴量をみやすくするため，縦軸のメモリの最大値を 1,000 としている．図 6 をみると，file:delDir など大きく差異がある特徴量もみられるが，200 未満の値を示す特徴量は赤と青の値が似ている傾向がみられる．よって一部の特徴量を除き，おおむね正しく予測できたといえる．また図 7 をみると，図 6 とは異なり，赤と青の特徴量の値の差異が大きくなっていることが分かる．正しく特徴量が予測できていないために，不正解であったと考えられる．

以上の結果より，正しく特徴量を予測できた検体は正しく分類でき，正しく特徴量を予測できなかった検体は正しく分類できない傾向にあることを確認した．この確認により，精度向上に正しい予測が寄与したことが分かった．

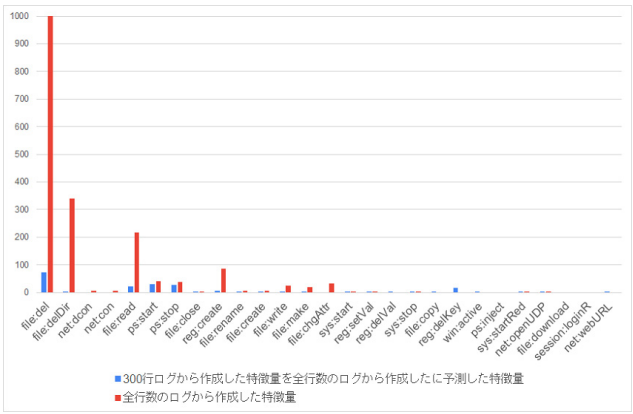


図 7 全行数のログから作成した特徴量を使用した分類で正解，300 行ログから作成した特徴量を全行数のログから作成した特徴量に予測した際の分類で不正解であった検体

Fig. 7 Correct answer in using the features of 300lines, in-Correct answer when predicted the features to ones of 300lines from ones of 20lines.

8. まとめと今後の課題

本論文では，できるだけ短い動的解析ログで高精度に最終進行度を推定するために，回帰を用いて長いログの特徴量を予測し，そして表層情報を組み合わせた最終進行度推定を提案した．実験の結果，回帰を行わず分類 1 のみで行った精度よりも，提案手法は最大で 2.5%向上した．また各検体のすべてのログで分類した場合と同程度の精度を得るのに必要なログの行数を，提案手法は 97.8%削減することに成功した．

今後はさらにデータ数を増やした実験を行い，提案手法の有効性の確認や，予測精度向上に努めたい．また現状は動的解析ログとして Mark II ログのみ有効性を確認したため，今後は他の動的解析ログに対して実験を行うことで有効性を確認したい．

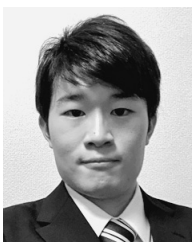
参考文献

[1] 竹下隆史ほか：マスタリング TCP/IP 入門編第 5 版，オーム社 (2018).
[2] Cichonski, P. et al.: Computer Security Incident Handling Guide, *NIST* (2012).
[3] 岡山あんほか：動的解析ログと表層情報を組み合わせたマルウェア感染活動の最終進行度推定手法，コンピュータセキュリティシンポジウム 2021 論文集，pp.1021–1028 (2021).
[4] Leyla, B. et al.: RiskTeller: Predicting the Risk of Cyber Incidents, *Proc. 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp.1299–1311 (2017).
[5] Kichang, K. et al.: Risk Assessment Scheme for Mobile Applications Based on Tree Boosting, *IEEE Access*, Vol.8, pp.48503–48514 (2020).

[6] 西野琢也ほか：テンソル分解に基づくグラフ分類による組織内ネットワーク攻撃活動検知，コンピュータセキュリティシンポジウム 2017 論文集 (2017).
[7] 矢野正太郎ほか：組織内ネットワーク攻撃進行度の自動推定技術の評価検証，第 80 回全国大会講演論文集，pp.453–454 (2018).
[8] Yuvraj, S.T. et al.: Real Time early Multi Stage Attack Detection, *2021 7th International Conference on Advanced Computing and Communication Systems* (2021).
[9] MITRE ATT&CK: ATT&CK, MITRE (online), available from ⟨https://attack.mitre.org/⟩ (accessed 2021-08-08).
[10] Nitesh, K. et al.: Malware Classification using Early Stage Behavioral Analysis, *2019 14th Asia Joint Conference on Information Security* (2019).
[11] Matilda, R. et al.: Early-Stage malware prediction sing recurrent neural networks, *Computers&Security*, Vol.77, pp.578–594 (2018).
[12] 朝倉紗斗至ほか：動的解析ログを用いた特徴量の予測によるマルウェアの早期機能推定に関する検討，コンピュータセキュリティシンポジウム 2020 論文集，pp.602–609 (2020).
[13] Sudhir, K. et al.: A Lifecycle Based Approach for Malware Analysis, *2014 Fourth International Conference on Communication System and Network Technologies* (2014).
[14] 寺田成吾ほか：通信挙動に基づくマルウェア種別分類手法，コンピュータセキュリティシンポジウム 2017 論文集 (2017).
[15] 寺田真敏ほか：マルウェア対策のための研究用データセット MWS Datasets〜コミュニティへの貢献とその課題〜，情報処理学会研究報告，Vol.2020-IFAT-139, No.8 (2020).
[16] Cuckoo: Cuckoo Sandbox 2.0.7, Cuckoo (online), available from ⟨https://cuckoosandbox.org/blog/207-interim-release⟩ (accessed 2021-08-08).
[17] GitHub: FFRI/FEXRD, Github (online), available from ⟨https://github.com/FFRI/FEXRD⟩ (accessed 2021-08-08).
[18] scikit learn: scikit-learn, scikit (online), available from ⟨https://scikit-learn.org/stable/⟩ (accessed 2021-08-08).
[19] 服部雄一ほか：確率統計入門，培風館 (2008).

岡山 あん

2021 年電気通信大学情報理工学域卒業．同年同大学大学院情報理工学研究科情報学専攻入学．マルウェア研究に従事．



朝倉 紗斗至

2020 年電気通信大学情報理工学域卒業．2022 年同大学院情報理工学研究科情報学専攻修士課程修了．在学中は動的解析ログを用いたマルウェアの早期目的推定に関する研究に従事．

中川 恒

2018 年に株式会社 FFRI に入社．入社以降，R&D 部門にてサイバーセキュリティの基礎研究に従事．現在は，macOS，Arm 版 Windows を対象とした脆弱性研究，および新規技術研究のディレクションを担当．2020 Black Hat EU Briefings 登壇，2021 CODE BLUE 登壇，2020，2021 セキュリティ・キャンプ全国大会講師．

押場 博光

2014 年に株式会社 FFRI に入社．自社製品の研究開発部門において自社製品のマルウェア検知ロジックの開発および，そのための情報収集・分析業務に従事．現在は新規事業創出を目指した新技術研究やそのディレクションを担当．2020 Black Hat EU Briefings 登壇，2020，2021 セキュリティ・キャンプ全国大会講師．

市野 将嗣

2003 年早稲田大学理工学部電子・情報通信学科卒業．2008 年同大学大学院理工学研究科博士課程修了．2007 年日本学術振興会特別研究員．2009 年早稲田大学大学院基幹理工学研究科研究助手．2010 年同大学メディアネットワークセンター助手．2011 年電気通信大学大学院情報理工学研究科助教．2016 年同大学院情報理工学研究科准教授．バイオメトリクス，ネットワークセキュリティに関する研究に従事．博士（工学）．電子情報通信学会各会員．