［サイバー・ウォーズ］

② 機械学習を用いたサイバーセキュリティ技術の発展

高橋健志　古本啓祐　韓　燦洙　情報通信研究機構

求められる自動化技術

増加するサイバースペース上の脅威に対応するためには，脆弱性管理やマルウェア分析をはじめ，さまざまなセキュリティオペレーションを実施する必要がある．しかしながらそれを十分に実現するのに必要な人的リソースは不足しており，その拡充およびオペレーションの効率化・自動化が必須である．前者については大学等教育関係機関が尽力しているが，後者については我々研究者が検討している領域である．ここでは，人のナレッジやノウハウをコンピュータに伝授し，徐々にコンピュータにより代替する手法を検討する必要があるが，機械学習はそれを実現する有効な手段であり，積極的に活用されるようになってきている．

機械学習は画像認識，テキスト分類，感情分析，音声認識，バーチャルパーソナルアシスタントなどの領域で広く活用されてきており，近年の発達は目覚ましいものがある．サイバーセキュリティ領域においても，スパムメールフィルタリング，フィッシング検知，フェイクニュース検知，フェイクレビュー検知，生体認証，マルウェア分析，自動運転セキュリティなど，さまざまな領域にてすでに活用されてきている．これは最近になって発生した事象ではなく，たとえばスパムメールフィルタリングについては 10 年以上前から検討が進められてきており，その当時から機械学習自体の弱点を突いた攻撃の可能性まで検討がなされてきている．とはいえ，最近になって機械学習を適用するのに必須となる計算機環境およびツール群の入手が容易になってきたことにより，より広くさまざまなケースに機械学習を活用する試みがなされるようになってきている．もはや機械学習を用いること自体には新規性はなく，機械学習は強固なサイバーセキュリティを実現するのに必須不可欠なツールとなっている．本稿では，機械学習がサイバーセキュリティ領域の研究開発でどのように活かされているのか，その概観をお伝えすべく，いくつかの領域での機械学習の活用事例を紹介し，その後に我々が実施している活動を紹介したい．

活用される機械学習技術

セキュリティオペレーションの自動化が扱う技術領域は広い．各種のオペレーションのすべての領域で機械学習の活用の可能性は検討されているだろう．とはいえ機械学習を用いる都合上，学習対象となるデータセットを準備可能な領域でしか研究開発を実施することはできない．データセット種別の観点から当該領域を大まかに整理すると，図 -1 に示す通り，通信トラフィック・ログの分析，ソフトウェアのバイナリやコードの分析，そしてテキスト情報の分析といった 3 つの領域に整理することができる．これらの領域では，そ

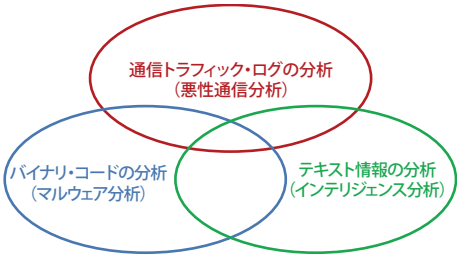

図 -1　主な分析対象領域

れぞれ悪性通信検知，マルウェア検知，インテリジェンス分析などの研究がなされている．実際には，サイバーセキュリティオペレーションの自動化を高い精度で実施するためには，これらの複数の領域にまたがって分析を実施する必要があることも多い．この整理を精緻に検討する意義はないものの，下記ではおおむねこの領域区分に沿っていくつかの領域をピックアップし，研究開発の現状を俯瞰したい．なお，関連する研究領域は広く，上記に限るものではない．たとえば，匿名化された投稿から投稿者を特定するプライバシ関連の技術や，機械学習自体のセキュリティについても多数の報告がなされているが，それらについては本稿では省略する．

通信トラフィック・ログを分析する

通信を分析する研究領域では，ネットワーク上の通信パケットそのものを分析する，もしくは通信のログを分析する等のものが存在する．通常の通信とは異なる通信や，マルウェア特有の通信パターンに類似するものを検知する技術など，さまざまなものが存在している．その中で筆者らが追跡している研究領域を下記に紹介する．

ネットワーク侵入検知の自動化

ネットワーク侵入検知システム（IDS）には大きくシグネチャ方式と異常検知方式が存在する．シグネチャ方式は誤検知率は低いが，シグネチャが登録されていない未知の攻撃（ゼロデイ攻撃）を検知することができない．シグネチャ方式は既知のシグネチャから攻撃パターンを分類または異常なデータから新しいシグネチャを抽出することに機械学習手法が用いられている．異常検知方式はゼロデイ攻撃を検知できる可能性があるものの，誤検知率が比較的高い．異常検知方式は異常なネットワークトラフィックを分類・クラスタリングする手法が用いられている．通常IDSは両方式を組み合わせたハイブリッド方式が用いられる．この分野は長期にわたり研究がなされている分野であるが，いまだに進展が報告されている．たとえば，トラフィックデータから特徴を抽出し，そこから改めて元データを生成する際に生じる再構成誤差に着目し，その値が大きい際に異常判定する技術などが提案されている．

悪性URLの検知・評価

フィッシングやドライブバイダウンロード攻撃を引き起こす悪性URLを特定すべく，異なるデータセットを用いたさまざまな方式が報告されている．いくつかのアプローチが存在するが，その1つは，ホスト名やプライマリドメイン，トップレベルドメインなど，URLの文字列自体に悪性サイトの特徴があると考えるものである．これらのものの中には，Whois情報やAS番号，ドメインの年齢など，ドメインに関する情報を利用するのも報告されている．別のアプローチとして，悪性URLに到達するまでのURL遷移を追跡し，その遷移から読み取れる情報，すなわち悪性URLまでのホップ数などを特徴として悪性URLがソーシャルエンジニアリング攻撃かドライブバイダウンロード攻撃のどちらに属するものかを判別することも可能になってきている．また，実際に訪問するページ自体を分析するアプローチも存在する．スタイルシートやJavaScript，ページのリンク構造など，さまざまな着眼点からの分析が報告されており，たとえばJavaScript分析の際の障壁となる難読化対策については，JavaScriptのコードをバイトコードレベルで分析することにより解決する技術などが報告されている．

ユーザの悪性サイトアクセス傾向分析

上述の悪性URL分析の研究の延長線上に位置づけられるものではあるが，目指しているものが悪性URLの検出ではなく，ユーザ自体の特性の分析になっている研究も存在する．たとえば，携帯電話網での通信ログから各ユーザのWebブラウジング記録に関する特徴量を抽出し，それに基づき同ユーザが近い将来に悪性URLを訪問する可能性があるかどうかを予測する手法なども報告されている．

暗号化通信の分析

通信を分析する技術領域では，暗号化通信の分析

も多数報告されている．セキュリティ対策を実施する際には，時に暗号化された通信やバイナリを分析する必要があり，暗号化された通信の中身を識別するのに機械学習を活用する研究もさまざまなものが報告されている．たとえば，Android および iOS アプリから発せられる暗号化された通信を分析することでアプリのフィンガープリントを抽出する技術が報告されている．さまざまなアプリやマルウェアが常に登場するため，教師あり学習を用いることによるスケーラビリティが課題となることが多いため，教師なし学習を用いた手法も検討されている．

バイナリ・コードを分析する

バイナリを分析する研究領域では，アプリの中からマルウェアを検知する技術や，その機能を分析する技術が検討されてきている．ここではそのうちのマルウェアの分類技術について概観を紹介する．マルウェアの機能とその潜在的な影響を分析するマルウェア解析技術は，セキュリティ分野における重要課題の１つである．マルウェア解析には大きく，サンドボックス上でマルウェアを実際に実行して調査する動的解析と，マルウェアを実行せずに調査する静的解析がある．これらの解析には従来，属人的な高い分析スキルが必要とされていたが，機械学習を用いてその解析を代替する技術の検討が進められている．

動的解析に基づくマルウェア検知

従来大量のマルウェア検体を自動的に検知・分類するのにシグネチャを用いていたが，そのシグネチャの準備には多大な工数を要するため，機械学習の活用によりシグネチャを用いない分析技術が検討されてきた．その特徴量には，逆アセンブルが不要な動的解析を用いるものが多数報告されており，システムコールやアプリケーションプログラミングインターフェース（API）などの呼び出し情報を特徴量として抽出する．マルウェア検知に加え，そのファミリーの分類なども実現可能である．

バイナリ分析に基づくマルウェア検知

近年は機械学習を用いた静的解析に関する報告が活発になっている．バイナリを分析する際にソースコードが手元にあれば，そのままそれを分析することが可能であるが，マルウェアを分析する際にはソースコードを入手できるケースは限定的である．そのため，分析を実施する前に，通常はバイナリを逆アセンブルもしくは逆コンパイルする．そこから得られたコードから特徴量を抽出し，分析を実施する．マルウェア検知という目的では，バイナリ間の類似度を評価することにより，マルウェアに近しいものか否かを判断する技術が存在するが，利用する特徴量，およびその類似度の指標などにもさまざまなものが報告されている．特徴量としては，バイナリから再構成した制御フローグラフや命令文そのもの，関数の引数の数や型など，さまざまなものが用いられており，それらは深層学習等の機械学習技術を用いてベクトル化され，そのベクトルを元に，バイナリ間の距離を評価する．これらの技術により，マルウェアの新種・亜種を効率的に検出することが可能となり，その結果に基づきマルウェアのシグネチャを自動で作成することで，より一層マルウェア解析の自動化が期待できる．

コード分析に基づくマルウェア検知

ソースコードが提供されている場合には，上述のような逆コンパイル処理は不要であり，より分析しやすいソースコードをそのまま分析することが通例である．また，ソースコードが提供されていなくとも，Android の APK や Java などのように比較的容易に人間が解読可能なコードへと変換可能なものも存在する．これらのコードを分析する際には，たとえば API コールや関数名を特徴量としたクラスタリングなどを実施することで，マルウェア検知やその機能分類などを実施することができる．

テキスト情報を分析する

テキスト情報を分析する研究領域では，SNS 等の投稿のセンチメンタル分析やフェイクニュース分析など，さまざまなものがあるが，ここではその中の脅威インテリジェンスに焦点を絞り，紹介する．脅威インテリ

ジェンスとは，アナリストなどにより生成された，セキュリティオペレーションに資する有益かつ重要な情報であり，通常，信頼性が高いと考えられる．脅威インテリジェンスとして収集された情報を，セキュリティインシデントの解析へ活用する動きが近年進んでいる．以下に，そのインテリジェンスを生成・分析する技術についていくつか紹介する．

脅威インテリジェンスの生成

各セキュリティベンダが定期的に発行しているセキュリティレポートや個人のブログ記事，Twitter などのSNS 上に投稿されているインシデント情報は，自組織内で発生したセキュリティインシデントの解析に有用である．しかし，これらの情報は自然言語で記述されており，独自のフォーマットで記載されていたり，独自のタグ付けがなされているケースがほとんどである．そのため，非構造化された脅威情報を解析に利用する場合，そのままではコンピュータへの侵入を示す痕跡情報などと自動的に突合させることや適切なラベルを自動付与することは難しい．これまではセキュリティオペレータが経験に基づいてセキュリティレポートなどから有益な情報を収集してきたが，機械学習技術を活用することにより，非構造化文書からのインテリジェンス抽出，ラベル付け，そして標準化された共通フォーマットにてレポートを自動生成する技術が検討されてきている．その際には，機械学習だけではなく，抽象的な概念モデルであるオントロジなどを活用するケースなども見受けられる．

情報のアノテーション

自然言語からインテリジェンスを生成する技術と近しいものに，情報のアノテーションを付与する研究も報告されている．これは，脆弱性や脅威に関する情報について，より詳しい情報を付加するものである．たとえば，インシデントレポートや脅威情報に対し，その内容がサイバーキルチェーンの中のどのフェーズに関するものなのかを分析し，その情報を付与するものが報告されている．サイバーキルチェーンは標的型攻撃などの攻撃活動を 7 つのフェーズに分解する考え方である．これは攻撃者の行動分析に役立てることが可能である．

ノイズ情報の検知・特定

多数のセキュリティ情報が存在する一方で，中にはノイズとなる情報が存在するケースも存在する．構造化された情報の中にもそのようなものが存在することもあり，それを特定する研究も報告されている．このようなノイズは，ヒューマンエラーに起因することも多く，ノイズの原因を特定し，修正することでより価値の高いインテリジェンスを生成することが可能となる．たとえば，自然言語情報から類推される脆弱なソフトウェアバージョン群と，実際にインテリジェンスに記載されている当該情報を比較し，一致しないものを自動検出する技術などが報告されている．上記のノイズに悪意があるケースとして，たとえばフェイクニュースなども存在しており，それを検知する技術も報告されている．セキュリティオペレーションで用いられるデータセットの中に悪意のあるユーザアカウントから投稿された情報が含まれている可能性も十分に考えられ，データセットの汚染を検知する技術についても検討が進められている．

データセットの準備

ここまで複数の領域をピックアップしてサイバーセキュリティ領域での機械学習技術の活用研究について紹介したが，本分野においては，研究データセットをいかにして確保するかが重要となる．研究を実施したくとも，その研究を実施するための分析対象となるデータセットを構築できなければ，機械学習は適用できない．そして，競争力の高い研究開発を実施するためには，質の高くユニークなデータセットと，信頼性および説得力のあるラベルを付与することが必要不可欠となる．

競争力のあるデータを収集する

大手検索エンジンや ISP，ソーシャルメディアサービ

スプロバイダなどは，膨大なデータを持っており，それらのデータを利用できることにより競争力の高い研究を実施することが可能となる．こういったデータにアクセスできない場合には，それらのデータと差別化のできるデータをいかにして収集・構築するかが，最終的な研究成果の質に大きく影響する．

筆者らの例では，我々はダークネット通信網を 10 年以上の期間観測しており，この期間の長さが競争力の源泉となっている．また，数千人の被験者を集め，ブラウザ上での Web アクセス履歴収集なども実施しており，大手検索エンジンから比べると規模は小さいものの，ユーザのブラウザ上での操作を追跡できる粒度での情報収集が，差別化要因になっている．

データをラベリングする

教師あり学習を実施する際には，データにラベルを付与する必要があるが，基準が明確でなく苦慮するケースが多い．たとえばマルウェア検知技術の場合には，あるバイナリがマルウェアであるか否かのラベル情報を付与することになるが，何をもってマルウェアとすべきかという基準は明確ではない．さまざまな工夫がなされているが，約 90 社のアンチウイルス製品の結果が収集できる VirusTotal の情報を元に，複数の情報源から多数決でラベル名を決定する手法なども提案されている．また，専門家によるラベル付けを実施するケースも多数報告されているが，通常，取り扱うデータのコンテクストでの専門家は機械学習の専門家ではない．そのため，機械学習の専門家ではない方々が効率的かつ効果的にラベル付けを実現可能にするためのインタフェースの研究なども報告されている．

我々の取り組み

日本では，2018 年度に官民研究開発投資拡大プログラム（PRISM）が創設されている．これは高い民間研究開発投資誘発効果が見込まれる領域に各府省庁の研究開発施策を誘導し，官民の研究開発投資の拡大，財政支出の効率化等を目指すものであり，情報通信研究機構は 2019 年度には九州大学，神戸大学，横浜国立大学，早稲田大学と連携し，この PRISM の AI 技術領域においてサイバー攻撃ハイブリッド高速分析プラットフォームの研究開発を実施してきた．我々はこれまでも機械学習を用いたサイバーセキュリティの研究開発を実施していたが，PRISM により方向性を持って技術を連携・発展させることにより，技術の飛躍的向上を実現すべく，研究活動を実施している．

図 -2 に，サイバー攻撃ハイブリッド高速分析プラットフォームの研究開発概要を示す．本研究開発では，インターネット上で新たなマルウェア活動の発生を自動的に瞬時に検知し，それに関連する情報を自動的に生成･抽出し，必要なエンティティに必要なセキュリティアラートを提供する技術の研究開発を実施している．分析対象となるデータとして，ダークネットトラフィック，ライブネットトラフィック，マルウェアサンプル，脅威インテリジェンスなどを主に利用している．それらのデータから特徴抽出・前処理を実施した上で，複数の分析を並行して実施し，その分析結果をリアルタイムに

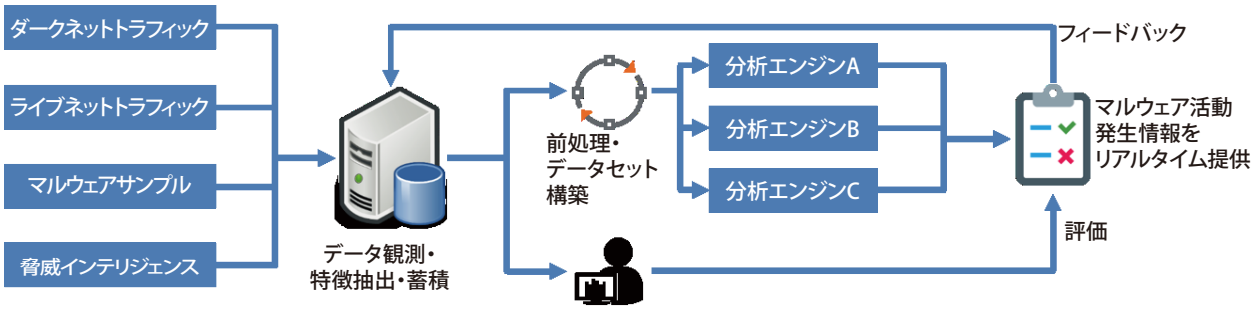

図 -2　サイバー攻撃ハイブリッド高速分析プラットフォームの研究開発概要

ユーザに提供する．また，アナリストによる評価を適宜実施し，その結果を反映した学習も可能な形になっている．分析対象領域としては，上述にある通信トラフィック・ログの分析，バイナリ・コードの分析，テキスト情報の分析の 3 領域すべてにまたがる研究開発を実施しているが，その中で今回の目的に合致した技術を絞って検討していく．

本研究開発ではさまざまな要素技術の研究を実施しているが，筆者らが手掛けているものをいくつか以下に紹介する．まず，我々は悪性の通信を検知・分析する研究として，マルウェア活動の発生検知，およびセキュリティアラートのスクリーニング技術の研究を実施している．前者ではインターネット上のダークネットと呼ばれる利用されていないアドレス空間に到着するパケットを分析することにより，マルウェアの活動が新たに発生するのを自動的に検知する技術を検討している．後者では，セキュリティアプライアンスが生成するセキュリティアラートの中で，特に重要性の高いものを自動的に抽出する技術を検討している．また，マルウェアを含む攻撃のツール自体を検出・分析する研究として，マルウェアの系統樹分析，および囮文書の評価の研究開発を実施している．前者では，マルウェアの機能を分類し，マルウェアの進化の過程が分かる系統樹の自動構築を検討している．後者では，標的型メールに添付されるファイルの内容の信憑性評価を実施している．

サイバー攻撃ハイブリッド高速分析プラットフォームを実現するには，上記を含む各種の要素技術を連携させる必要がある．本研究開発では，これらの要素技術の効果的な連携手法も検討しているが，その際には実際のオペレータの利用を想定した検討を進めている．

今後の研究開発の方向性

機械学習により，これまで実現し得なかったセキュリティオペレーションの自動化が実現し始めている．特に，人手により分類することが処理時間的に非現実的であったものが自動化されるメリットは大きい．スパムメールの検知や通常とは異なる通信の検知，マルウェア検体候補のスクリーニングなど，機械学習により効率化を実感できているオペレーションも存在する．一方で，機械学習だけですべてのセキュリティ課題を解決することはできない．実際，現時点でのセキュリティ学会での発表を俯瞰すると，機械学習に依存しない発表も多数存在する．また，機械学習の限界を示唆する論文も提出されてきている．

これまでの数年間は機械学習の適応可能性を広く探してきた時代であると感じているが，これからはその可能性の模索を継続して実施するとともに，より深く分析するための機械学習技術の活用の検討が求められると考えている．より深く分析するためには，領域が細分化されがちな研究業界において，積極的にその領域の壁を乗り越える努力が必要不可欠である．そして，セキュリティオペレーションの現場にあるノウハウをより色濃く吸収し，機械学習などにより再現することにより，地道に技術を成長させていく必要がある．攻撃者は技術の発展を待ってくれるわけではないため，従来のヒューリスティックベースの技術の発展と機械学習技術を用いたセキュリティ技術の発展を並行して実施し，融合していくことで，スピード感のある研究開発をしていきたい．

（2020 年 4 月 6 日受付）

高橋健志　takeshi_takahashi@nict.go.jp
2005 年早稲田大学大学院博士課程修了．タンペレ工科大学，（株）ローランド・ベルガーを経て，現在，情報通信研究機構に勤務．機械学習技術を用いたセキュリティ業務の自動化技術の研究開発に注力．CISSP，情報処理安全確保支援士．

古本啓祐（正会員）　k.furumoto@nict.go.jp
2018 年神戸大学大学院博士課程後期課程修了．現在，情報通信研究機構に勤務．機械学習を用いた脅威インテリジェンス分析およびトラフィック解析に関する研究に従事．

韓　燦洙　han@nict.go.jp
2018 年九州大学大学院修士課程修了．現在，同大学博士後期課程在学，および情報通信技術機構に勤務．機械学習を用いたマルウェアおよびトラフィック解析に関する研究に従事．