サイバー脅威インテリジェンス生成自動化

角丸 貴洋　高橋 航　勝瀬 陸　シラクザーノ ジュセッペ　サンヴィート ダヴィデ　ビフィルコ ロベルト

要 旨

NECでは、サイバーセキュリティリスクの早期発見に向け、インテリジェンスアナリストが日々サイバーセキュリティ情報を収集し、蓄積や分析を行っています。しかし、収集対象が政治・経済・社会・技術動向などサイバー攻撃以外にも広がっており、いかに幅広い領域を収集対象としつつも情報源を的確に絞り込み分析を行うかが課題でした。分析を高精度かつ迅速に行うために、生成AIを活用しサイバー脅威インテリジェンスの生成自動化に取り組んでいます。本稿では、NECのサイバー脅威インテリジェンスの取り組みと課題、そして、開発中のサイバー脅威情報の抽出・要約パイプラインと、サイバー脅威関連情報の探索・分析パイプラインを紹介します。

KeyWords　生成AI／サイバー脅威インテリジェンス／戦略インテリジェンス／情報抽出技術／統合分析

1. はじめに

サイバー脅威インテリジェンスとは、サイバー脅威に関する情報を収集し分析することを言い、そのプロセスだけでなく、成果物としての情報も指します[1)]。一般的なインテリジェンスには、戦術インテリジェンス、運用インテリジェンス、戦略インテリジェンスの３種類があり[2)]、利用する人の立場や目的がそれぞれ異なります。

特に、戦略インテリジェンスでは、経営層がセキュリティリスクに対して行う意思決定をサポートするため、組織を取り巻く脅威についてサイバー攻撃のみならず、政治・経済・社会・技術などの外部環境も踏まえながら、さまざまな角度からサイバー脅威を分析することが求められます。

主に戦略インテリジェンスに関する情報は、レポート、ブログ、記事、ニュース、SNSなどさまざまな書式で記載されており、また、発信者ごとに異なる表現が使われることもあるため、情報の収集や整理は容易ではありません。このため、インテリジェンスアナリストは高度な知識と経験が求められてきました。しかし、これからの情報収集では、より幅広く収集し目的とする情報を探し出すことが求められ、熟練のアナリストだけに頼らない、かつ過度に人に依存した情報収集からの脱却が必要です。

また、収集・分析の要求を受けてから数日以内、場合によっては数時間以内に報告が求められることもあります。経営層の期待に応えるには、正確で、目線が合っていて、意思決定ができる情報をタイムリーに提供する必要があるため、迅速性が重要となります。

このような背景から、アナリストの業務負荷を低減し、分析能力を向上させるインテリジェンス生成プロセスの自動化・省力化が必要になります。

2. 戦略インテリジェンスの取り組みと課題

戦略インテリジェンスの利用者が求めるのは、今起こっているのがどんな問題で、他の類似する組織の状況はどうなっているのか、求められる意思決定は何かといった情報です。インテリジェンスアナリストは必要なファクトを集め、「想像力を働かせながら」分析して仮説を生み出すことで経営層の期待に応えます。

意思決定者からの要求を受けて情報を収集及び分析し、行動を起こすために必要な情報を生成する一連の流れはインテリジェンスサイクルと呼ばれています[3)]。いくつかのバリエーションがありますが、ここでは要件定義・収集・処理・分析/作成・配布・フィードバックのステップで

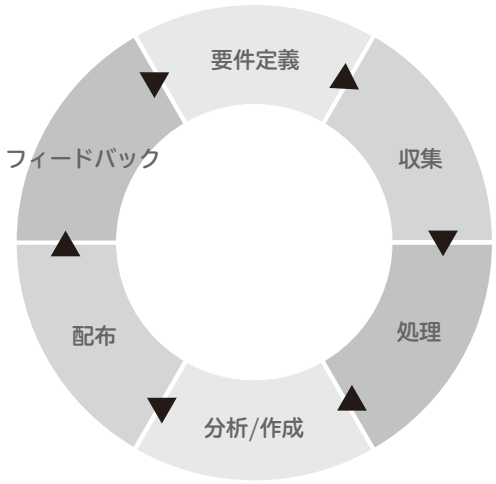

図1　インテリジェンスサイクル

サイクルを表します（**図1**）。

一方、要求を受けて一から情報を収集し、それを整理・分類したのでは、更に分析をしたうえで意思決定に資する情報に的確に仕上げることを考えると、期待されるタイミングには到底間に合いません。日頃から情報を収集・蓄積し、すぐに使える状態にしておくことが重要となります。また、顕在化する脅威や脅威の変化にいち早く気付くためにも、日々情報に目を配っておくことが鍵となります。

そのため、NECでは、日々インテリジェンスアナリストが収集し整理しているサイバー脅威に関する情報をデータベースに蓄積しています。しかし、収集対象の範囲が政治・経済・社会・技術動向などサイバー脅威情報にとどまらなくなってきていることから、収集から整理、蓄積までをできる限り人手が掛らないようにし、幅広い領域を収集対象としていくという課題がありました。

更に、サイバー脅威の調査においては膨大な量の情報源から関係する情報を探し読み込んだうえで、必要なファクトを集め分析をすることになります。しかし、どこに必要とする情報があるのかを探すのに苦労したり、キーワードレベルでは合致していても求める情報とは違うため使えず無駄が発生したりするため、必要とする情報源を的確に絞り込むという課題がありました。

これらの課題を解決するために、NECが開発している生成AIを活用したサイバー脅威情報の抽出・要約パイプラインと、サイバー脅威関連情報の統合分析パイプラインを紹介します。

3. 生成AIを活用したサイバー脅威インテリジェンス生成自動化の実現方針

戦略インテリジェンスの重要性を考えると、次の課題を解決することが最も重要です。

（1）生成AIモデルが生成したコンテンツには誤った情報が含まれている可能性があること（幻覚）

（2）生成AIモデルは、入力として提供された長い文書に現れる重要な情報を無視する可能性があること（長い文脈）

課題（1）に対する解決策として、生成AIを推論ツールとして活用します。具体的には、生成AIモデルに外部から与えたデータセットのみを用いて回答するよう指示します。更に、生成AIモデルが提供する回答の精度を確認するための検証プロセスを導入します。このプロセスには、外部の情報源による検証や、AIモデルとの対話を通じた自己評価や推論手順の実施など、さまざまなアプローチが含まれます。

課題（2）に対する解決策として、長い文書を生成AIに送信する前に、前処理ステップを導入します。この段階では、元の文章を分析し、タスクと関連性の高い内容だけを選択してテキストから情報を抽出します。このフィルタリングプロセスは、インテリジェンスアナリストの要求に応じて調整され、自動化されるタスクに応じて変更することが可能です。

4. サイバー脅威情報の抽出・要約パイプライン

4.1 パイプラインの概要

インテリジェンスアナリストは、レポート、ブログ、記事、ニュース、SNSなどの多様な非構造化情報源から情報を収集・整理する際、手作業で大量の文書を読み解き、情報をデータベースで適切かつアクセスしやすい構造化フォーマットに変換しなければなりません。

NECのサイバー脅威情報の抽出・要約パイプラインは、これらのタスクを自動化し、効率化を図ります。このパイプラインは、3つの要素によって構成されています（**図2**）。

ダウンローダー/パーサーは、HTMLなどの形式で表現されたさまざまな自然言語のレポートをテキストのみの形式に変換します。特定のよく知られた情報源については、フォーマットを処理するプラグインを作成しておきます。このように変換されたテキストドキュメントが、この後に

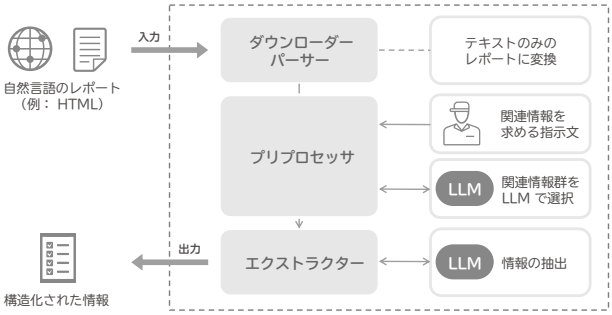

図２　サイバー脅威情報の抽出・要約パイプライン

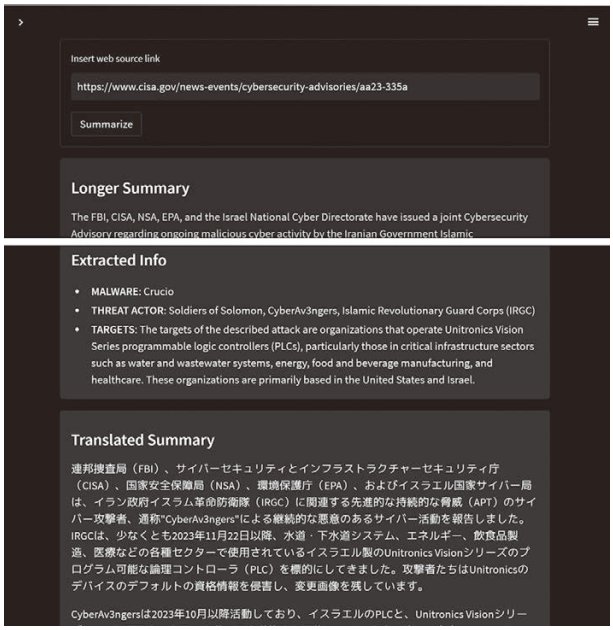

図３　情報の抽出と要約の活用例

生成AIで用いられる唯一のデータセットとなります。

プリプロセッサは、関連性のある情報を特定するために、反復的なフィルタリングを実行します。まず、入力テキストは短い段落に分割され、各段落は前の段落と数文ずつ重なるようにすることで文脈が失われないように調整されます。その後、新しく形成された段落から関連性のある情報を選択するために、生成AIが利用されます。

インテリジェンスアナリストが求めるような関連する情報の種類を、自然言語で記述できる指示文によって簡単に設定できます。例えば、「攻撃の経済的影響に関する情報を特定してください」といった指示文です。元の文章から、タスクと関連性が高い文章だけを選択して、それらが関連情報群としてまとめられます。

最後に、エクストラクターは、前のステップで作成された関連情報群としてまとめられたテキストを入力として受け取ると、生成AIを活用して情報を抽出し、指定されたデータベースで使用される構造化フォーマットに沿った形で出力します。これにより、抽出された情報が格納されます。

4.2 パイプラインの特徴

サイバー脅威情報の抽出・要約パイプラインには、2つの重要な特徴があります。

1つ目の特徴は、関連する情報を識別し、生成AIを利用して詳細に着目した抽出が可能であることです。実際、脅威レポートなどを手作業で分析する際、インテリジェンスアナリストは、レポートの視点を維持したまま、情報を抽出する際に何を省略し、何を含めるかを決定しなければなりません。この決定には通常、信頼性レベルやレポートに記載された情報の詳細性に関する考慮が含まれます。NECの生成AI技術は、アナリスト同様の分析と関連文章の選択を自動的に実行するよう簡単に指示できます。

2つ目の特徴は、要求されたタスクや出力先の要件に基づいて、抽出された情報のフォーマットを素早く調整して出力することができることです。生成AIへの指示文を調整することによって、文章中から特定の情報を抽出する場合であっても、要約であっても可能となります。

4.3 活用例

このパイプラインは、一連の情報ソースを自動的に監視し、構造化された情報を抽出するために使用されます。データベースの格納形式に合わせて、文章中から特定の情報を抽出、要約、イベントの分類が行われます。これにより、経験の浅いアナリストが脅威情報の収集と要約に要する時間を、約2時間から1時間へと約50%短縮できました（**図3**）。

5. サイバー脅威関連情報の取得・分析パイプライン

5.1 パイプラインの概要

インテリジェンスアナリストの仕事は、さまざまな情報源からの情報を処理し、適切なイベントを抽出し、異なる文書やツールからのイベントやデータを関連付けることです。サイバー脅威関連情報の取得・分析パイプラインは、

図4　サイバー脅威関連情報の取得・分析パイプライン

アナリストが実施するこのプロセスを支援し、作業を可能な限り自動化するために設計されています。

このパイプラインは、3つの異なる要素によって構成されています（**図4**）。

探索及び取得モジュールは、自然言語を入力としてインテリジェンス分析クエリを受け付けます。例えば、「特定の脅威アクターに関連するすべての攻撃について教えてください」、または「このファイルハッシュは脅威アクターXに関連していますか?」などです。最初のクエリをもとに、検索拡張生成（RAG、Retrieval Augmented Generation）の仕組みを用いて生成AIが外部の情報を参照できるようにします。

この際に使用される外部の情報としてさまざまな脅威データベースが使用でき、例えば、独自の脅威データベース、知識のつながりをグラフ構造で表した知識グラフ、ベクトル形式でインデックス化したベクトルストア、信頼できるWeb上の公開情報が含まれます。

これらの文書は次のモジュール（関連性の検証と関連箇所の抽出）に入力され、最初のクエリをもとに、検索された文書が実際にクエリに関連しているかどうかを検証します。関連性が確認できた場合、各文書からの情報は、文書の元々の視点を保持しつつ、最初のクエリに関連する情報を網羅した簡潔な要約として抽出されます。

最後に、関連性が確認され、要約された文書が分析モジュールに渡されます。このモジュールでは、関連する要約にある情報を結び付け、最初のクエリに対する回答を生成する推論者として、生成AIが活用されます。この段階で、推論された分析結果を更に確認するために外部ツールを呼び出すこともできます。例えば、マルウェアのハッシュを分析するサービスや、脅威データベースを使用した検証ができます。

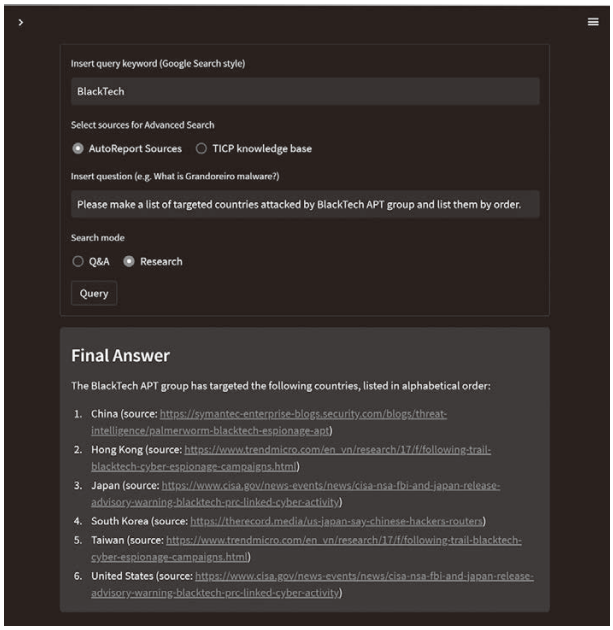

図5　情報の取得と分析の活用例

5.2 パイプラインの特徴

このパイプラインの大きな特徴は、さまざまな脅威データベースや複数の文書から膨大な数の関連情報を検索・処理することにより、信頼性の高い情報ソースを横断して文書分析が可能になることです。分析クエリとドキュメントの関連性を特定し、その観点を維持しながら内容を要約することで、精度の高い回答を提供することができます。更に、外部のツールや情報源を利用して提供された回答を検証することで、精度が更に向上します。

5.3 活用例

現在、サイバー脅威関連情報の検索・分析パイプラインは社内で検証が行われています。インテリジェンスアナリストは、Webインタフェースを通じて生成AIが組み込まれたシステムと対話することで、自動的に取得された情報と分析結果を得ることができます（**図5**）。これまでの検証から、より精度の高い情報が提供されるとともに、インテリジェンス業務の迅速化に寄与することが期待できます。

6. むすび

NECでは、サイバーセキュリティリスクを早期に発見するために、インテリジェンスアナリストが日々サイバーセキュリティ情報を収集し、情報の蓄積や分析を行っています。更なる情報源の拡大や分析能力の強化に向けて、生成AIを活用した自動化・効率化に取り組んでいます。

本稿では、生成AIの幻覚問題と長い文脈の課題を解決するサイバー脅威情報の抽出・要約パイプラインと、サイバー脅威関連情報の取得・分析パイプラインを紹介しました。一部の効果は確認できていますが、未検証の部分については引き続き検証を進めていきます。

NECで培ってきた分析ノウハウを更に組み込み、より柔軟で多様な脅威分析に対応できるように技術開発を進めていく予定です。

NECは、サイバーインテリジェンスへの取り組みを通じて、社会課題の解決と公平・効率という社会価値の創造に貢献します。

参考文献

1) ISO/IEC 27002:2022：Information security, cybersecurity and privacy protection - Information security controls
https://www.iso.org/standard/75652.html
2) アメリカ統合参謀本部：Doctrine for Intelligence Support to Joint Operations, 2000.3
https://www.hsdl.org/c/abstract/?docid=3735
3) アメリカ合衆国国家情報長官：U.S. National Intelligence An Overview 2011
https://www.dni.gov/files/documents/IC_Consumers_Guide_2011.pdf

執筆者プロフィール

角丸 貴洋
サイバーセキュリティ戦略統括部
グループ長
CISSP

高橋 航
サイバーセキュリティ戦略統括部

勝瀬 陸
サイバーセキュリティ戦略統括部

シラクザーノ ジュセッペ
NEC Laboratories Europe
Principal Research Scientist

サンヴィート ダヴィデ
NEC Laboratories Europe
Senior Research Scientist

ビフィルコ ロベルト
NEC Laboratories Europe
Senior Manager

NEC 技報のご案内

NEC技報の論文をご覧いただきありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報（日本語）　NEC Technical Journal（英語）

Vol.75 No.2　ビジネスの常識を変える生成AI特集
～社会実装に向けた取り組みと、それを支える生成AI技術～

ビジネスの常識を変える生成AI特集によせて
生成AI技術への取り組み ～基盤から応用、ルール作りまで～

◇ 特集論文

急速に広まる生成AIの市場適用
NECにおける生成AIの取り組みについて
電子カルテと医療文書の作成支援による医師業務効率化
映像×LLMによる報告書作成業務の自動化
映像分析と生成AIによるリアル世界の行動理解
サイバー脅威インテリジェンス生成自動化
生成AIの社内活用を進めるNEC Generative AI Service（NGS）
ソフトウェア・システム開発への生成AIの活用
LLMとMIで革新する素材開発プラットフォーム
LLMと画像分析を活用した被災状況の把握

生成AIの可能性を高める基盤技術
日本語性能に優れたNECの大規模言語モデル（LLM）
NECの生成AIを支える国内企業で最大規模のAIスーパーコンピュータ
より安全な大規模言語モデル（LLM）を目指して
データを秘匿したまま連携できる連合学習技術とLLMへの適用可能性
大規模言語モデル（LLM）によるFew-shotクラスタリングの可能性
オープンドメイン常識推論のための知識拡張型プロンプト学習
AI連携とオーケストレーションのための基盤ビジョンLLM
クエリを考慮した新規手法により関連する企業データを減らしLLM APIの使用コストを最適化

AI技術が社会へ浸透するために
AI標準化・ルールメイクに関する動向とNECの取り組み
人権尊重に向けたNECのAIガバナンスの取り組み
RCModelを用いたAIリスクマネジメントのための人材育成事例

◇ NEC Information
2023年度C&C賞表彰式典開催

Vol.75 No.2
（2024年3月）

特集TOP