4ZA-01

機械学習モデルの継続的更新のための脅威情報のアノテーション

箕浦 翔悟† 毛利 公美† 白石 善明‡

†岐阜大学工学部電気電子・情報工学科 ‡神戸大学大学院工学研究科／ATR

1. はじめに

サイバーセキュリティ分野においても機械学習を適用する試み[1][2]が盛んである．機械学習モデルの精度は様々な要因で決まるが，その中でも学習用データセットの良し悪しが大きな影響を与える．データの持つ意味が時間経過とともに変化するような運用環境では，構築されたモデルは作成時が最高の精度で，以降時間とともに精度が落ちていく．機械学習モデルの継続的な更新が変化に対応するための解決策であるが，そのための教師データを更新し続けることは，データの入手が容易であったとしてもラベル付けをする作業にコストがかかる．また，教師データの質の向上のためには一人より複数人で行うことが望ましい．本研究では文献[3][4]などの脅威情報を活用する研究開発のためのデータセットの構築と精度維持・向上を目的として，教師データの量を増やしながら質の向上を図る取り組みを支援するツールを提案する．

2. 脅威情報のアノテーション

2.1. コンセプトドリフト

モデル構築の当初はカテゴリーA に分類されていても，ある時点以降はカテゴリーB に分類されるべきことがあるように，データの内容はシステム運用の経過とともに変化することがある．特にセキュリティ関連では新しいマルウェアやサイバー攻撃等，新出の用語がそれに当てはまる．例えば，2017 年頃から活動しているマルウェアの名称“WannaCry”は，その登場以前はただの動詞に過ぎなかった．このように時間経過に伴いデータが変化していく現象をコンセプトドリフトという．

コンセプトドリフトによりモデルの劣化が懸念される．図 1 に示すように劣化を防ぐために常に新しい教師データを投入しモデルを更新することが望まれる．そこで本稿では脅威情報の機械学習モデルを継続的に更新するための教師データの作成を支援するアノテーションツールを提案する．

2.2. 要件定義

コンセプトドリフトに対応した機械学習モデルの更新のために以下のような流れのアノテーションを考える．まず，Web 上から脅威情報に関連するデータの収集を行い，その中から教師データとしてふさわしいデータを選別する．続いて，それらの選別されたデータに対して複数人でエンティティアノテーションを行う．このようなツールに求められる要件を以下に与える．

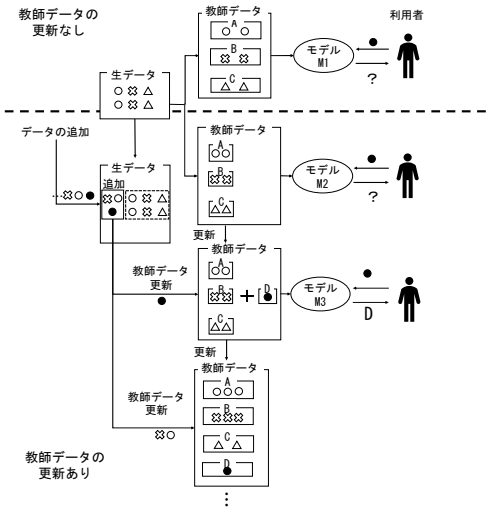

図 1 更新される機械学習モデルと継続的なアノテーションによる教師データの作成

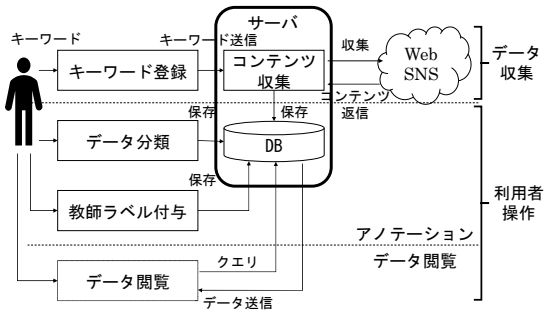

図 2 提案するアノテーションツールのシステム構成と機能配置

要件 1：データの自動収集・保存ができる
自組織が必要とする情報に関連するデータを自動で収集し，保存することが求められる．

要件 2：アノテーション対象データを選別できる
収集されたデータに対し，利用者は教師データとしてふさわしいものを選別する．

要件 3：複数人で作業ができる
教師データの質の向上のために複数人で作業を行えることが求められる．

要件 4：分散型の作業ができる
任意の端末から同じ操作方法で共同作業できるようにすることが望ましい．

3. 脅威情報のアノテーションツール

提案するツールのシステム構成を図 2 に示す．ツールは主に“**キーワード登録機能**”，“**コンテンツ収集機能**”，“**データ分類機能**”，“**教師ラベル付与機能**”，“**データ閲覧機能**”の機能から構成され，動作は次のようになる．

Annotation Tool for Continuous Update of Machine Learning Model

† Shogo MINOURA, Masami MOHRI・Gifu University

‡ Yoshiaki SHIRAISHI ・ Kobe University / Advanced Telecommunication Research Institute International

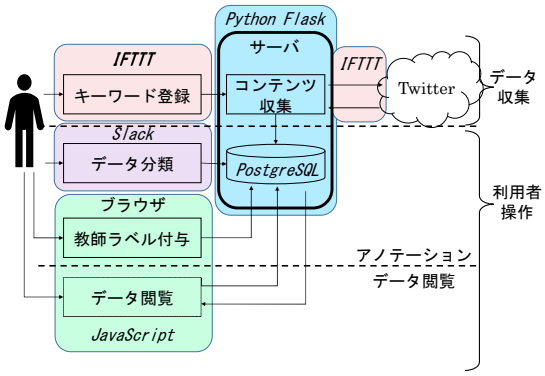

図 3　実装環境

図 4　データ閲覧画面

図 5　アノテーション画面

キーワード登録：利用者は収集したい情報に関するキーワードを登録する．【キーワード登録機能による：要件 1 に対応】

コンテンツ収集：システムは登録されているキーワードをもとにインターネットで検索して脅威情報を収集する．【コンテンツ収集機能による：要件 1 に対応】．

データ分類：収集されたコンテンツに対して，利用者はアノテーションするのにふさわしいものを分類し，その結果を記録する．【データ分類機能による：要件 2 および要件 4 に対応】

教師ラベル付与：利用者は分類されたコンテンツに対してエンティティアノテーション（ラベル付け）を行う．教師ラベル付与の結果は複数人に共有される．【教師ラベル付与機能による：要件 3 および要件 4 に対応】

データ閲覧：利用者は自動収集されたコンテンツ，アノテーション対象の分類結果，アノテーション結果を閲覧することができる．【データ閲覧機能による：要件 3 および要件 4 に対応】

4.　実装

サーバサイドは Python の Flask フレームワーク，ブラウザサイドは JavaScript で記述する．図 3 は，コンテンツ収集先を Twitter とし，キーワードを IFTTT に登録し，Slack に自動投稿されたツイートを Slack 上でアノテーション対象かどうか分類する実装例である．

ブラウザでのデータ閲覧画面およびアノテーション結果を図 4，図 5 に示す．図 4 の画面上でコンテンツの内容，アノテーション対象に分類されたか，アノテーション済みかを確認できる．編集ボタンをクリックすると図 5 の編集画面に遷移する．テキストをドラッグして選択，テキスト上部のラベルをクリックするとテキストにラベルを付与することができる．保存ボタンをクリックすると DB に保存した後に図 4 のデータ閲覧画面に戻る．

5.　まとめ

時間経過とともにデータの意味が変化し機械学習モデルの劣化が懸念される環境において，教師データの追加・更新という課題に対して，本稿ではその作業を支援するツールを提案した．

表 1 に既存のアノテーションツール brat [5]，doccano [6]との違いをまとめる．本提案は，継続的な更新という観点から，データの収集からアノテーションまで可能な限り滞りなくできるようにしたことで，要件 1 および要件 2 で差異を持つものである．しかしながら，高度なアノテーションを可能とする既存ツールとの連携は今後の検討課題の一つである．

表 1　アノテーションツールの比較

	brat [5]	doccano [6]	本提案
要件 1	−	−	○
要件 2	−	−	○
要件 3	○	○	○
要件 4	○	○	○
言語	Python 2	Python 3	Python 3

参考文献

[1] Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., and Wang, C., “Machine Learning and Deep Learning Methods for Cybersecurity”, IEEE Access, vol.6, pp.35365-35381 (2018).

[2] 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).

[3] Zhu, Z. and Dumitras, T., “ChainSmith: Automatically Learning the Semantics of Malicious Campaigns by Mining Threat Intelligence Reports”, 2018 IEEE European Symposium on Security and Privacy (EuroS&P), pp.458-472 (2018).

[4] Ito, D., Nomura, K., Kamizono, M., Shiraishi, Y., Takano Y., Mohri, M., Morii, M., “Modeling Attack Activity for Integrated Analysis of Threat Information”, IEICE Trans. on Information and Systems, vol.E101-D, no.11, pp.2658-2664 (2018).

[5] brat rapid annotation tool, http://brat.nlplab.org/index.html

[6] doccano, https://github.com/doccano/doccano