SCIS 2022 2022 Symposium on
Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 — 21, 2022
The Institute of Electronics,
Information and Communication Engineers

文書類似性モデル評価手法による潜在意味解析に基づくセキュリティレポート検索の評価

Evaluation of Security Report Retrieval Based on Latent Semantic Analysis Using Document Similarity Model Evaluation Method

添田 綾香† 長澤 龍成† 白石 善明†
冨田 裕涼†† 箕浦 翔悟†† 毛利 公美†† 森井 昌克†
Ayaka SOEDA Ryusei NAGASAWA Yoshiaki SHIRAISHI
Yusuke TOMITA Shogo MINOURA Masami MOHRI Masakatu MORII

キーワード 脅威情報，非構造化文書，トピックモデル，潜在意味解析，関連文書検索

サイバー攻撃による組織の被害を最小限にするために，セキュリティ管理者は現在起こっているインシデントに関する情報を収集し，迅速に対応しなければならない．汎用的な検索エンジンでは大量の情報が出力され，有用な情報を選別するのに時間と労力を要する．我々は図 1 に示すシステムのログに残った IoC 情報を用いて，幅広い脅威情報を収集できるセキュリティレポート検索システムを提案している[1]．

本システムではトピックモデルの一種である LDA (Latent Dirichlet Allocation)[2]を用いて[3]の手法で付与された特徴ベクトルを用いて検索を行う．利用者は初めに IoC 情報などのキーワードをクエリとした全文検索を行う．検索サーバは検索結果の文書に付いている特徴ベクトルを用いて関連文書検索を行い，全文検索と関連文書検索の結果を出力する．その検索結果から利用者が注目するレポートを選んで関連文書検索をさらに行うことができ，セキュリティ情報を話題に広がりを持たせて抽出する．

本論文では，文書類似性モデル評価手法[4]に基づいて提案システムの特徴ベクトルの作成手法がセキュリティ情報検索の望ましい結果を出力するかを評価している．LDA，Word2vec，fastText のそれぞれで特徴ベクトルを作成した検索システムを用意し，応用先の一つである NICT のサイバー攻撃ハイブリッド分析プラットフォーム[5]での使用を想定した検索クエリを用いて比較したところ，文献[3]の手法で作成した特徴ベクトルがより良い結果を抽出できることを確認している．

† 神戸大学大学院工学研究科電気電子工学専攻，神戸市
Kobe University, 1–1 Rokkodai-cho, Nada-ku, Kobe-shi,657–8501 Japan
†† 岐阜大学工学部電気電子・情報工学科，岐阜市
Gifu University, 1–1 Yanagido, Gifu-shi, 501–1193 Japan

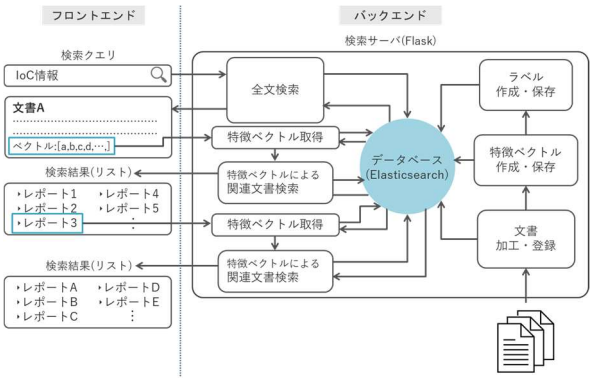

図 1 セキュリティレポート検索システムの構成と機能.

Figure 1 Construction and functions of proposed security report retrieval

参考文献

[1] 添田綾香，長澤龍成，長田侑樹，白石善明，冨田裕涼，箕浦翔悟，毛利公美，森井昌克，“サイバー攻撃分析のためのセキュリティ情報検索システム”，マルチメディア、分散、協調とモバイル (DICOMO2021) シンポジウム，pp.874-882，2021 年 7 月.

[2] D.M. Blei, A.Y. Ng and M.I. Jordan, “Latent Dirichlet Allocation,” Journal of Machine Learning Research, pp.993-1022, March 2003.

[3] 長田侑樹，瀧田愼 ，古本啓祐，白石善明，高橋健志，毛利公美，高野泰洋，森井昌克，“セキュリティ情報検索のためのトピックモデルによるマルチラベリング，” 暗号と情報セキュリティシンポジウム (SCIS2021)，2C3-3，Jan. 2021.

[4] K. Krstovski, D. Smith, M. Kurtz, “Automatic Construction of Evaluation Sets and Evaluation of Document Similarity Models in Large Scholarly Retrieval Systems”, Scholarly Big Data: AI Perspectives, Challenges, and Ideas Workshop, AAAI, 2016.

[5] T. Takahashi, Y. Umemura, C. Han, T. Ban, K. Furumoto, O. Nakamura, K. Yoshioka, J. Takeuchi, N. Murata, Y. Shiraishi, "Designing Comprehensive Cyber Threat Analysis Platform: Can We Orchestrate Analysis Engines?", Proc. of IEEE International Conference on Pervasive Computing and Communications (PerCom), March 2021.