

テクニカルノート

# 単語のトピック固有度を用いた脆弱性記述に基づく脆弱性特性の自動評価

中川 舜太<sup>1</sup> 白石 善明<sup>1,a)</sup> 古本 啓祐<sup>2</sup> 毛利 公美<sup>3</sup> 森井 昌克<sup>1</sup>

受付日 2021年2月17日, 採録日 2021年6月7日

**概要:** ソフトウェアの脆弱性は組織に甚大な損失をもたらす危険性がある。セキュリティ担当者は、脆弱性を狙った攻撃のリスクを最小限にとどめるために、脆弱性特性を参考にしてより深刻な脆弱性から対策を立てることになる。脆弱性特性の評価は人手によるコストや遅延が懸念されており、自動で評価を行うことが求められている。本論文では、脆弱性記述をもとに脆弱性特性を評価する既存研究に単語のトピック固有度を加えた手法について提案している。比較実験により、単語のトピック固有度を用いた提案手法が既存手法よりも多くの脆弱性特性項目で予測精度が上回っていることを確認している。

**キーワード:** CVSS, CVE, 脆弱性評価, 機械学習

## Automated Software Vulnerability Assessment Based on Vulnerability Descriptions Using Topic Specificity of Words

SHUNTA NAKAGAWA<sup>1</sup> YOSHIAKI SHIRAISHI<sup>1,a)</sup> KEISUKE FURUMOTO<sup>2</sup>  
MASAMI MOHRI<sup>3</sup> MASAKATSU MORII<sup>1</sup>

Received: February 17, 2021, Accepted: June 7, 2021

**Abstract:** Software vulnerabilities pose a risk of bringing serious losses to organizations. In order to minimize the risk of vulnerabilities being exploited, security professionals should address the more serious vulnerabilities first. For this purpose, it is necessary to assess the vulnerability. In this paper, we propose new method to assess vulnerabilities based on vulnerability descriptions by adding topic specificity of words to existing research. Through comparative experiments, we confirmed that our method using the topic specificity of words outperforms the existing studies in many items.

**Keywords:** CVSS, CVE, vulnerability assessment, machine learning

### 1. はじめに

ソフトウェアにおける脆弱性は、個人や組織に大きな損失をもたらす危険性がある。組織においてシステム管理者やセキュリティ担当者は、脆弱性を狙った攻撃のリスクを最小限にとどめるために、深刻な脆弱性から対策を立てる必要があり、優先順位を決めなければならない。

そのために、脆弱性スコアリングフレームワークである CVSS (Common Vulnerability Scoring System) [1] が用いられる。CVSS は評価項目に答えることで脆弱性の深刻度が算出される。人手による脆弱性の評価は、コストが大きく、脆弱性評価の遅延を起こす。また、評価者がセキュリティの実務に精通していなければ、ミスを起こす可能性もある。そのため、脆弱性の自動評価システムが求められる。

これまでに、脆弱性を自動評価する手法が提案されている [2], [3], [4]。Le ら [4] は、脆弱性の概要文に基づいて自動で CVSS 評価する手法を提案している。本論文では、比較的単純なモデルで CVSS の 6 つの基本評価基準のそれぞれと深刻度を推定している Le らの手法に対して単語のトピック固有度を導入する手法を提案し、その予測精度につ

<sup>1</sup> 神戸大学  
Kobe University, Kobe, Hyogo 657-8501, Japan

<sup>2</sup> 情報通信研究機構  
National Institute of Information and Communications  
Technology, Koganei, Tokyo 184-8795, Japan

<sup>3</sup> 岐阜大学  
Gifu University, Gifu 501-1193, Japan

a) zenmei@port.kobe-u.ac.jp

いて比較する。

## 2. 脆弱性識別子と脆弱性スコアリングフレームワーク

本研究では、脆弱性識別子である CVE (Common Vulnerabilities and Exposures) [5], 脆弱性スコアリングフレームワークである CVSS (Common Vulnerability Scoring System) [1] を用いる。

### 2.1 CVE

CVE は、個別製品中の脆弱性を対象として、MITRE 社が採番している識別子である。CVE には主に CVE-ID, CVE Description (以下、脆弱性記述と呼ぶ)、参照 URL の 3 つの情報が含まれている。本研究では、脆弱性の自動評価のための入力として、脆弱性記述を用いる。

### 2.2 CVSS

CVSS は、情報システムの脆弱性に対する汎用的な評価手法である。CVSS は、基本評価基準、現状評価基準、環境評価基準の 3 つの基準からなる。本研究では、時間経過によって変化せず、どの利用者でも活用できる脆弱性そのものの特性を評価する基本評価基準の評価結果を自動で出力することを目的としている。CVSS v2 の基本評価基準は Access Vector, Access Complexity, Authentication, Confidentiality Impact, Integrity Impact, Availability Impact の 6 つの評価項目からなる。評価項目は、それぞれ 3 種類の評価選択肢があり、評価者は 1 つの脆弱性につきそれぞれの項目から 1 つずつ選択する。それを既定の計算式に入力すると、CVSS 基本値 (0.0~10.0) が算出される。CVSS 基本値に基づいて、3 段階 (Low: 0.1~3.9, Medium: 4.0~6.9, High: 7.0~10.0) の深刻度が決定される。

## 3. 関連研究

### 3.1 脆弱性記述に基づく脆弱性特性評価

Le ら [4] は、脆弱性記述を用いて CVSS 基本評価基準と深刻度について推定する手法を提案している。文字レベルの n-gram (Bag of Words) と単語レベルの n-gram を組み合わせた特徴ベクトルを作成し、機械学習モデルの入力としている。また、自然言語処理設定や機械学習モデルの選択を行う際に、ランダムにデータを分割する一般的な交差検証ではなく、CVE の発行日時をもとにデータの分割を行う時間ベースの交差検証手法を提案している。これは、コンセプトドリフトが起こる脆弱性記述において、予測精度が下がらないようにすることを目的としている。

### 3.2 非構造化セキュリティ文書の構造化

Zhu ら [7] は、非構造化セキュリティ文書から IOC (Indicator of Compromise) を抽出し、攻撃を複数段階に分割

するサイバーキルチェーンフェーズを推定する研究を行っている。Zhu らは、文書内から重要な単語を抽出するためにトピックに固有な単語の抽出を行っている。単語のトピック固有度については、次の評価式で計算される。

$$S(w) = \max_{t \in T} \frac{p(w|t)}{p(w)}$$

ただし、 $w$  は評価対象の単語、 $t$  はラベル (トピック)、 $T$  はラベル  $t$  の集合、 $p(w)$  は文書全体における単語  $w$  の出現率、 $p(w|t)$  はラベル  $t$  の中での単語  $w$  の出現率である。

Zhu らは、 $S(w)$  がしきい値を超えた場合、その単語を重要な単語と見なし、機械学習モデルの入力に用いている。

## 4. 提案手法：単語のトピック固有度を用いた特徴抽出

脆弱性記述を入力として、CVSS v2 の 6 つの基本評価基準と深刻度の推定を行う。

### 4.1 テキスト前処理

テキスト前処理については、Le ら [4] と同様の処理を行う。句読点 (punctuation) やストップワード (stop words) を除去する。ファイル名 (例: input.c) などの用語を分離しないように保つために、句読点の後ろに空白を含む句読点のみを除去する。ストップワードリストとして、nltk [8] と scikit-learn [9] のストップワードリストを用いる。また、すべての文字を小文字に変換する。さらに、文字列は異なるが、同じ意味を表す単語 (例: “allow” と “allows”) を同様に扱うために、ステミングを行う。ここでは、nltk ライブラリの Porter Stemmer アルゴリズム [10] を用いる。

### 4.2 特徴抽出

手法 1 は、Le らの手法の単語 n-gram・文字 n-gram 作成時に単語のトピック固有度を考慮したものである。手法 2 は、Word2Vec を用いて単語や文字 n-gram をベクトル化し、Zhu ら [7] が用いた単語のトピック固有度の評価式  $S(w)$  を掛け合わせた単語レベル・文字レベルそれぞれの平均ベクトルの結合ベクトルを特徴として用いる。この節では、それぞれの手法について説明する。

#### 4.2.1 手法 1: Le らの手法に単語のトピック固有度を考慮した n-gram の作成

手法 1 の概要について、図 1 に示す。

訓練フェーズでは、まず、脆弱性記述から単語 n-gram ( $w$ ) と文字 n-gram ( $c$ ) を抽出する。このとき、Le らは総文書数のうちその単語が使われた文書の割合が下回ると除外される最小出現頻度を単語 n-gram では 0.001、文字 n-gram では 0.1 としていた。手法 1 では、1 つのラベルにおいてのみよく現れる、ラベルに特徴的な単語を除外しないようにするために、ラベルごとに最小出現頻度を単語レベルでは 0.001、文字レベルでは 0.1 として単語 n-gram と文字

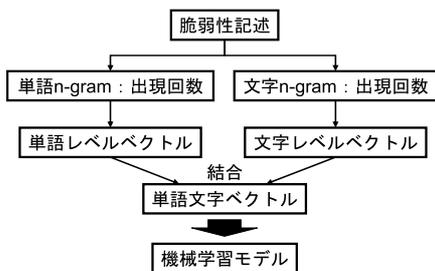


図 1 手法 1 の概要

Fig. 1 Outline of the method 1.

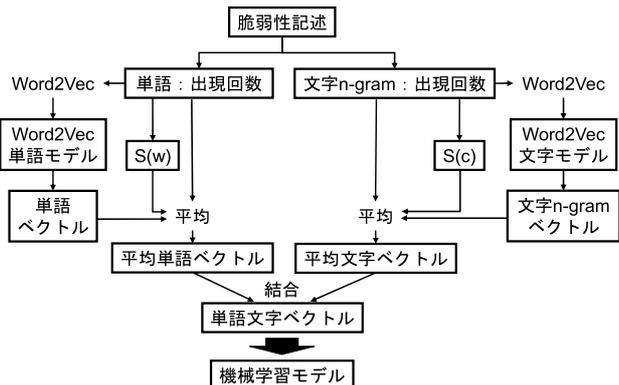


図 2 手法 2 の概要

Fig. 2 Outline of the method 2.

n-gram のリストを作成したうえで、全文書に対してそれらを語彙とする単語 n-gram と文字 n-gram を作成する。ここではそれらを単語レベルベクトルと文字レベルベクトルと呼ぶ。以降は、Le らと同様に、単語レベルベクトルと文字レベルベクトルを結合し、単語文字ベクトルを作成する。これを機械学習モデルに入力して訓練を行う。

予測フェーズでは、それぞれの脆弱性記述において単語 n-gram、文字 n-gram を作成し、これらを結合して単語文字ベクトルを作成する。これを訓練済み機械学習モデルに入力して予測を行う。

#### 4.2.2 手法 2 : Le らの手法に単語のトピック固有度と単語の分散表現を用いた特徴抽出

手法 2 の概要について、図 2 に示す。

訓練フェーズでは、まず、手法 1 と同様にラベルごとに脆弱性記述から単語と文字 n-gram を抽出する。ただし、のちに平均をとるため、文字レベルの最小出現頻度は 0.01 としている。単語と文字 n-gram について、これらの出現回数をもとに単語・文字 n-gram のトピック固有度  $S(w_i)$ ,  $S(c_j)$  を計算する。次に、脆弱性記述を単語ごとに区切ったリストを入力として、Gensim [11] の Word2Vec を用いて単語の分散表現を訓練し、Word2Vec 単語モデルを作成する。文字レベルでは、文書を n-gram の  $n$  の数だけずらしてリストを作成し、これらを Word2Vec の入力として訓練を行い、Word2Vec 文字モデルを作成する。たとえば、“hello world” という文があった場合、3-gram では、[“hel”,

“lo”, “wor”, “ld”], [“he”, “llo”, “ wo”, “rld”], [“h”, “ell”, “o w”, “orl”, “d”] という 3 種類のリストを作成する。本研究では、Word2Vec の学習時のパラメータは文字レベル・単語レベルともに埋め込み次元数は 300、ウィンドウサイズは 5 とする。

単語のトピック固有度一覧を  $S_w = (S(w_1), S(w_2), \dots, S(w_p))$ , Word2Vec 単語モデルによって出力された単語ベクトル一覧を  $V_w = (v_{1w}, v_{2w}, \dots, v_{pw})$ , 単語の出現回数一覧を  $F_w = (f_{1w}, f_{2w}, \dots, f_{pw})$ , 文字 n-gram のトピック固有度一覧を  $S_c = (S(c_1), S(c_2), \dots, S(c_q))$ , Word2Vec 文字モデルによって出力された文字 n-gram ベクトル一覧を  $V_c = (v_{1c}, v_{2c}, \dots, v_{qc})$ , 文字 n-gram の出現回数一覧を  $F_c = (f_{1c}, f_{2c}, \dots, f_{qc})$  とする。平均単語ベクトル  $w_{average}$ , 平均文字ベクトル  $c_{average}$  は以下の式で表される。

$$w_{average} = \frac{S_w \circ F_w V_w}{\sum S(w_i)n_i}$$

$$c_{average} = \frac{S_c \circ F_c V_c}{\sum S(c_j)n_j}$$

求めた平均単語ベクトルと平均文字ベクトルを結合し、単語文字ベクトルを作成する。これを機械学習モデルに入力して訓練を行う。

予測フェーズでは、それぞれの脆弱性記述において単語や文字 n-gram の出現回数と訓練済みの Word2Vec モデルの出力ベクトルと  $S(w)$ ,  $S(c)$  を用いて平均単語ベクトルと平均文字ベクトルを作成し、結合することで、単語文字ベクトルを作成する。これを機械学習モデルに入力して予測を行う。

## 5. 実験

手法 1, 手法 2 と Le らの手法について比較を行い、脆弱性評価における単語のトピック固有度や分散表現の有効性について検証する。

### 5.1 実験方法

データセットとして、NVD (National Vulnerability Database) [12] に掲載されている 1988 年から 2018 年の 113,292 件の脆弱性を用いている。脆弱性記述に \*\* REJECT \*\* の記載があるものや CVSS v2 の値を持っていないものについては除去し、105,124 件の脆弱性を扱う。2015 年以前のデータを訓練データ (76,241 件), 2016 年以降のデータをテストデータ (28,883 件) として用いた。

実験で用いた機械学習モデルと自然言語処理設定は、表 1 に示す。LGBM は LightGBM, XGB は XGBoost, LR は Logistic Regression を表している。これは、Le らが実験で示した最適なモデルや自然言語処理設定に基づいている。ただし、手法 2 では tf-idf を用いず、単語 n-gram の  $n$  は 1 としている。

表 2 実験結果

Table 2 Experimental results.

基本評価基準	手法 1			手法 2			Le らの手法		
	Accuracy	Macro F-Score	Weighted F-Score	Accuracy	Macro F-Score	Weighted F-Score	Accuracy	Macro F-Score	Weighted F-Score
Confidentiality	<u>0.736</u>	<u>0.726</u>	<u>0.737</u>	0.709	0.701	0.711	0.729	0.719	0.730
Integrity	<u>0.763</u>	<u>0.750</u>	<u>0.764</u>	0.727	0.715	0.729	<u>0.763</u>	0.749	<u>0.764</u>
Availability	0.710	0.707	0.709	<u>0.719</u>	<u>0.715</u>	<u>0.720</u>	0.712	0.711	0.712
Access Vector	<u>0.914</u>	0.537	0.899	<u>0.914</u>	<u>0.553</u>	<u>0.903</u>	<u>0.914</u>	0.540	0.901
Access Complexity	<u>0.704</u>	<u>0.469</u>	<u>0.674</u>	<u>0.704</u>	<u>0.470</u>	<u>0.677</u>	0.698	0.460	0.668
Authentication	<u>0.876</u>	<u>0.447</u>	<u>0.847</u>	0.864	0.426	0.832	0.875	0.442	0.844
Severity	0.666	0.589	0.664	<u>0.685</u>	0.566	<u>0.673</u>	0.670	<u>0.596</u>	0.668

表 1 機械学習モデルと自然言語処理

Table 1 Machine learning models and NLP settings.

基本評価基準	機械学習モデル	単語 n-gram	文字 n-gram	tf-idf
Confidentiality	LGBM	1	2-6	-
Integrity	XGB	1-3	2-6	-
Availability	LGBM	1	2-6	-
Access Vector	XGB	1-3	2-6	✓
Access Complexity	LGBM	1	2-6	-
Authentication	LR	1-2	2-6	-
Severity	LGBM	1-4	2-6	-

5.2 実験結果

実験結果を表 2 に示す。それぞれの評価値の中で最も高い数値を下線で示す。

手法 1 は、多くの項目で Le らの手法よりも高くなっていることが分かる。また、低くなっている項目においてもあまり低下していないことが分かる。すなわち、単語のトピック固有度を考慮することで予測精度が向上している。

また、手法 2 は、Availability, Access Vector, Access Complexity, Severity では Le らの手法より予測精度が高くなっている。一方で、Confidentiality や Integrity については Le らの手法と比べて予測精度が低下している。Han ら [2] や Gong ら [3] は、本手法 2 と同様に脆弱性記述のベクトル化に Word2Vec を用いているが、分類に CNN (Convolutional Neural Network) や LSTM (Long short-term memory) を使用している。これらは文中の単語の並びを考慮した学習が可能な機械学習モデルである。手法 2 では単語の分散表現を使用しているが、文中の単語の語順を考慮していない。これらのことから、単語の語順を考慮した機械学習モデルにおいて、単語の分散表現は有効であると考えられる。語順を考慮した機械学習モデルにおいて単語の分散表現の有効性の検証を行うことが今後の課題である。

6. まとめ

本研究では、脆弱性記述に基づいて、単語のトピック固有度を用いて脆弱性特性を推定する手法を提案した。実験の結果、単語のトピック固有度を用いることで、既存研究よりも予測精度が向上することを確認できた。語順を考慮する機械学習モデルでは単語の分散表現の活用が有用であることを示唆する結果が得られたため、そのようなモデルによる評価を行うことが今後の課題である。

謝辞 本研究は、総務省の「電波資源拡大のための研究開発 (JPJ000254)」における委託研究「電波の有効利用のための IoT マルウェア無害化/無機能化技術等に関する研究開発」による成果が含まれます。ここに感謝の意を表します。

参考文献

- [1] FIRST: Common vulnerability scoring system (CVSS), available from (<https://www.first.org/cvss/>) (accessed 2021-02-04).
- [2] Han, Z., Li, X., Xing, Z., Liu, H. and Feng, Z.: Learning to Predict Severity of Software Vulnerability Using Only Vulnerability Description, *Proc. 2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pp.125–136, IEEE (2017).
- [3] Gong, X., Xing, Z., Li, X., Feng, Z. and Han, Z.: Joint Prediction of Multiple Vulnerability Characteristics Through Multi-Task Learning, *Proc. 2019 24th International Conference on Engineering of Complex Computer Systems (ICECCS)*, pp.31–40, IEEE (2019).
- [4] Le, T.H.M., Sabir, B. and Babar, M.A.: Automated software vulnerability assessment with concept drift, *Proc. 16th International Conference on Mining Software Repositories*, pp.371–382, IEEE (2019).
- [5] MITRE: Common vulnerabilities and exposures (CVE), available from (<https://cve.mitre.org/>) (accessed 2021-02-04).
- [6] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).
- [7] Zhu, Z. and Dumitras, T.: ChainSmith: Automatically Learning the Semantics of Malicious Campaigns by Min-

ing Threat Intelligence Reports, *Proc. 2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp.458–472, IEEE (2018).

- [8] Bird, S., Loper, E. and Klein, E.: *Natural Language Processing with Python*, O'Reilly Media Inc (2009).
- [9] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E.: Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, Vol.12, pp.2825–2830 (2011).
- [10] Porter, M.F.: An algorithm for suffix stripping, *Program*, Vol.14, pp.130–137 (1980).
- [11] Rehurek, R. and Sojka, P.: Software Framework for Topic Modelling with Large Corpora, *Proc. LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp.45–50, ELRA (2010).
- [12] NIST: National Vulnerability Database (NVD), available from (<https://nvd.nist.gov/>) (accessed 2021-02-04).



中川 舜太 (学生会員)

2019年神戸大学工学部電気電子工学科卒業。2021年同大学大学院博士課程前期課程修了。FIT2018 ヤングリサーチ賞。機械学習に基づくサイバー攻撃分析の研究に従事。



白石 善明 (正会員)

1995年愛媛大学工学部情報工学科卒業。1997年同大学大学院博士前期課程修了。2000年徳島大学大学院博士後期課程修了。博士(工学)。2002年近畿大学理工学部情報学科講師。2006年名古屋工業大学大学院情報工学専攻助教授。2013年神戸大学大学院電気電子工学専攻准教授。2017年同大学数理・データサイエンスセンター兼任。2018年より(株)国際電気通信基礎技術研究所客員研究員。情報セキュリティ、コンピュータネットワーク、教育支援、知識流通支援、機械学習に基づくサイバー攻撃分析等の研究・教育に従事。2002年電子情報通信学会オフィスシステム研究賞、2003年暗号と情報セキュリティシンポジウム(SCIS) 20周年記念賞、2006年SCIS論文賞。2007, 2008, 2011, 2013, 2021年DICOMO優秀論文賞、2012年電子情報通信学会ライフインテリジェンスとオフィス情報システム研究会功労賞、2015年本会高度交通システム研究会優秀論文賞、2017年電子情報通信学会関西支部活動功労賞、2021年電子情報通信学会教育功労賞。2017年、2018年電子情報通信学会情報通信システムセキュリティ研究専門委員会委員長。電子情報通信学会シニア会員。



古本 啓祐 (正会員)

2013年神戸大学工学部電気電子工学科卒業。2014年同大学大学院博士課程前期課程修了。2018年同大学院博士課程後期課程修了。同年情報通信研究機構サイバーセキュリティ研究所。サイバーセキュリティと機械学習に関する研究開発に従事。



毛利 公美

1993年愛媛大学工学部情報工学科卒業。1995年同大学大学院工学研究科情報工学専攻博士前期課程修了。2002年博士(工学)(徳島大学)。1995年香川短期大学助手。1998年徳島大学工学部知能情報工学科助手、2003年同講師。2007年岐阜大学総合情報メディアセンター准教授、2017年同大学工学部電気電子・情報工学科准教授。ネットワークセキュリティ、符号・暗号理論、コンピュータネットワーク等の研究・教育に従事。電子情報通信学会シニア会員。



森井 昌克

1989年大阪大学大学院工学研究科通信工学専攻博士課程修了。工学博士。同年京都工繊大工芸学部電子情報工学科助手、1990年愛媛大学工学部情報工学科講師、1992年同助教授、1995年徳島大学工学部知能情報工学科教授、2005年神戸大学工学部電気電子工学科教授。現在、同大学大学院工学研究科教授。主に代数的符号理論、離散数学、デジタル信号処理アルゴリズム、情報セキュリティおよびコンピュータネットワーク等の研究・教育に従事。電子情報通信学会情報セキュリティ専門委員会、同ライフインテリジェンスとオフィス情報システム専門委員会、同情報通信システムセキュリティ専門委員会各委員長を歴任。2007年度電子情報通信学会功労賞、2010年度FIT船井ベストペーパー賞、2013年度電子情報通信学会論文賞、2020年情報セキュリティ文化賞。電子情報通信学会フェロー。