Next-Generation Security Entity Linkage:

Harnessing the Power of Knowledge Graphs and Large Language Models

Daniel Alfasi
Dept. of Computer Science
Reichman University
Israel
daniel.alfasi@post.runi.ac.il

Tal Shapira
Dept. of Computer Science
Reichman University
Israel
talshapirala@gmail.com

Anat Bremler-Barr
Dept. of Computer Science
Tel-Aviv University
Israel
anatbr@tauex.tau.ac.il

ABSTRACT

With the continuous increase in reported Common Vulnerabilities and Exposures (CVEs), security teams are overwhelmed by vast amounts of data, which are often analyzed manually, leading to a slow and inefficient process. To address cybersecurity threats effectively, it is essential to establish connections across multiple security entity databases, including CVEs, Common Weakness Enumeration (CWEs), and Common Attack Pattern Enumeration and Classification (CAPECs). In this study, we introduce a new approach that leverages the RotatE [4] knowledge graph embedding model, initialized with embeddings from Ada language model developed by OpenAI [3]. Additionally, we extend this approach by initializing the embeddings for the relations.

CCS CONCEPTS

• **Security and privacy** → **Vulnerability management**.

KEYWORDS

CVE, CWE, CAPEC, Knowledge Graph Embedding

1 SOLUTION

We created a knowledge graph with CVE, CWE, and CAPEC entities and nine types of relations between them. Our approach is based on knowledge graph representation learning, and we employed a multi-modal model. Specifically, we employ RotatE model, which has shown effectiveness in modeling complex relationships, and we further improve its performance by using pre-trained embeddings for both entities and relations. These embeddings are obtained by feeding the entity descriptions into the Ada language model that capture the semantic information of each entity.

Previous research has focused on using knowledge graphs to link security entities, such as the approach taken by Han et al. [5]. Other works have utilized both entity descriptions and graph structure, as in the case of Xing et al. [6], who proposed a text-enhanced Graph Attention Network that utilizes a word2vec [1] model to extract textual features from entity descriptions. However, these studies have not provided benchmarking datasets. To address this issue, we have created two datasets: one consisting of 4,096 Linux CVEs from 1999-2020, obtained from the MITRE CVE database [2], and a new dataset containing 16,044 CVEs from the RedHat Security Database.

SYSTOR '23, June 5–7, 2023, Haifa, Israel

ACM ISBN 978-1-4503-9962-3/23/06.
https://doi.org/10.1145/3579370.3594759

2 RESULTS

In contrast to previous works, we assess our method using an Inductive Link Prediction protocol, which handles unseen entities during training. Our approach is improving the Mean reciprocal rank (MRR), Hit@10, Hit@5, and Hit@1 by 11%, 5%, 7%, and 15%, respectively as shown in Table 1. Using MRR with Hit@ offers a more comprehensive evaluation, as MRR considers ranking of correct answers while Hit@ only checks if the correct answer is in the top k results.

Table 1: Evaluation of Linux CVEs dataset.

Model	MRR	Hit@10	Hit@5	Hit@1
Xing et al. [6]	0.49	0.65	0.59	0.4
Our approach	0.6	0.7	0.66	0.55

Furthermore, when comparing our model initialization to other methods such as Word2Vec, our results highlight the efficacy of our approach, as shown in Table 2.

Table 2: RotatE evaluation with different initializations.

Model	MRR	Hit@10	Hit@5	Hit@1
RotatE+W2V	0.54	0.64	0.59	0.5
RotatE+Ada	0.6	0.7	0.66	0.55

3 ACKNOWLEDGMENT

This work was partially supported by Red Hat.

REFERENCES

[1] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *1st International Conference on Learning Representations, ICLR 2013.* http://arxiv.org/abs/1301.3781
[2] MITRE. 1999. CVE. https://cve.mitre.org
[3] OpenAI. 2022. Ada Embedding. https://openai.com/blog/new-and-improved-embedding-model
[4] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *CoRR* abs/1902.10197 (2019). arXiv:1902.10197
[5] Hongbo Xiao, Zhenchang Xing, Xiaohong Li, and Hao Guo. 2019. Embedding and Predicting Software Security Entity Relationships: A Knowledge Graph Based Approach. In *Neural Information Processing*. Springer International Publishing, Cham, 50–63.
[6] Liu Yuan, Yude Bai, Zhenchang Xing, Sen Chen, Xiaohong Li, and Zhidong Deng. 2021. Predicting Entity Relations across Different Security Databases by Using Graph Attention Network. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC).* 834–843.