Chapter 11
Detecting Phishing URLs With Word Embedding and Deep Learning

Ali Selamat
https://orcid.org/0000-0001-9746-8459
Universiti Teknologi Malaysia, Malaysia & Hradec Kralove University, Czech Republic

Nguyet Quang Do
Universiti Teknologi Malaysia, Malaysia

Ondrej Krejcar
https://orcid.org/0000-0002-5992-2574
Hradec Kralove University, Czech Republic

ABSTRACT

The past decade has witnessed the rapid development of natural language processing and machine learning in the phishing detection domain. However, there needs to be more research on word embedding and deep learning for malicious URL classification. Inspired to solve this problem, this chapter aims to examine the application of word embedding and deep learning in extracting features from website URLs. To achieve this, several word embedding techniques, such as Keras, Word2Vec, GloVe, and FastText, were used to learn feature representations of webpage URLs. The obtained feature vectors were fed into a deep-learning model based on CNN-BiGRU for extraction and classification. Two different datasets were used to conduct numerous experiments, while various metrics were utilized to evaluate the phishing detection model's performance. The obtained findings indicated that when combined with deep learning, Keras outperformed other text embedding methods and achieved the best results across all evaluation metrics on both datasets.

DOI: 10.4018/978-1-6684-7684-0.ch011

INTRODUCTION

Phishing is currently a major area of interest within the field of cyber security. In recent years, there have been numerous efforts to mitigate phishing attacks and protect end users from losing their private and sensitive information to cybercriminals. Especially, the past decade has witnessed the rapid development of natural language processing (NLP) and machine learning (ML) in many phishing detection-related tasks (Bharadwaj et al., 2022; Tajaddodianfar et al., 2020; Vinayakumar et al., 2018; Yuan et al., 2018). Phishing detection is usually divided into three categories: malicious URL classification, phishing website detection, and phishing email detection. Malicious URL classification comprises related studies solely focusing on the detection of phishing attacks using URL-based features (T. Feng & Yue, 2020; Huang et al., 2019). Meanwhile, phishing website detection makes use of various features extracted from web pages to classify malicious and benign websites (J. Feng et al., 2020; Le-Nguyen et al., 2021). Phishing email detection regards emails as the medium to conduct phishing activities and extracts features from the email's header and body for classification (Hasan et al., 2021). Even though these three approaches use different types of features, these attributes can be extracted manually or automatically using numerous feature representation techniques and various learning algorithms.

NLP and ML have been widely used in phishing website and email detection to represent and extract features from the content of web pages and emails. However, the extraction of content-based features is time-consuming and computationally expensive (Ya et al., 2019). As a result, researchers and security experts have shifted their attention to phishing detection based on only URL features. Yet, much of the research on phishing URL detection up to now has focused more on word embedding and traditional ML (Bharadwaj et al., 2022; Yuan et al., 2018). On the one hand, conventional ML techniques require manual feature engineering. On the other hand, they cannot handle a substantial amount of data, resulting in a deficiency in detection accuracy (Bello et al., 2021). In addition, URL structure is different compared to website and email text. URL sometimes contains meaningless words and more information can be found at the character level. Nevertheless, the existing character embedding method disregards the relationships between characters and fails to capture meaningful information in long sequences. Whereas, the word-level embedding techniques can discover the semantic and syntactic similarities among words (Le et al., 2018). Still, there has been little research on word embedding with deep learning to identify malicious URLs.

Motivated to solve these problems, this chapter aims to investigate the application of word embedding and deep learning (DL) in extracting features from website URLs. First, word-level embedding can discover the semantic meaning and syntactic structure within URL sequences. Second, DL can prevent hand-crafted feature engineering and third-party dependency. Third, the extraction of URL-based features can reduce computational complexity. To achieve this, website URLs are used as inputs and pre-processed using several word embedding techniques (Keras, Word2Vec, GloVe, and FastText). Next, the obtained feature representations are fed into DL layers consisting of CNN and BiGRU for feature extraction and classification. Finally, website URLs are identified as malicious or benign based on the probability calculated by the Sigmoid function in the output layer. The main objectives of this chapter are as follows:

- To conduct a comparative analysis using various word embedding techniques to obtain feature representations from website URLs.
- To propose a DL-based phishing detection model using CNN-BiGRU to combine their complementary effects and improve the overall performance accuracy.