Fuzzy Expert System to Detect Phishing in Websites

Dániel MÁCSAI
Ismael RUIZ GARCIA
Mauro VÁZQUEZ CHAS

Master in Artificial Intelligence

UNIVERSITAT ROVIRA i VIRGILI

Planning and Approximate Reasoning
Delivery 3

15th December 2024

Contents

1 Introduction

For this work, we

2 Task 1

To design the fuzzy expert system to detect phishing websites, we consulted [1]. In this paper, they list 87 posible features (boolean, floats and integers) that could matter in the detection of phishing websites.

2.1 Chosen Features

Chosen variables

We selected features from each of the three categories presented in the paper: URL-based features, content features and external features.

URL-based features:

- Phish Hints (51) (int) Phishing URLs use sensitive words to gain trust on visited web pages. The number of such words in URLs is considered as phishing indicato

- Domain Age (83) (int)

Content features:

- Ratio external hyperlinks (59) (float)

External features:

- Google Index (86) (boolean) Phishing websites live for short times and are often accessible through direct links sent to users in emails, they do not need to be indexed by Google. Web pages not indexed by Google are supposed phishing

- Page rank (87) (int) Phishing web pages are not very popular, hence, they suppose to save low page ranks compared with legitimate web pages. We use Openpagerank to get the value of this feature **In the end we chose Google Pagerank that is 0-10**

2.2 Fuzzy Sets

For the output variable, you must define four fuzzy categories: safe, weakly suspicious, strongly suspicions and phishing. In addition to the category class, we want to get a numerical score in the range of 0..100

2.2.1 Abnormal Subdomains

References

[1] Abdelhakim Hannousse and Salima Yahiouche. "Towards benchmark datasets for machine learning based website phishing detection: An experimental study". In: *Engineering Applications of Artificial Intelligence* 104 (2021), p. 104347. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2021.104347. URL: https://www.sciencedirect.com/science/article/pii/S0952197621001950.