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. The proposed features are divided into three categories: URL-based features, content features and external features. From this proposed variables, we selected 5 features that we consider relevant for the detection of phishing websites.

2.1 Chosen Features

2.1.1 URL-based features:

Phish Hints

- **Description:** Number of words in the URL that are typical of phishing websites
- **Integer**. Number 51 in the paper

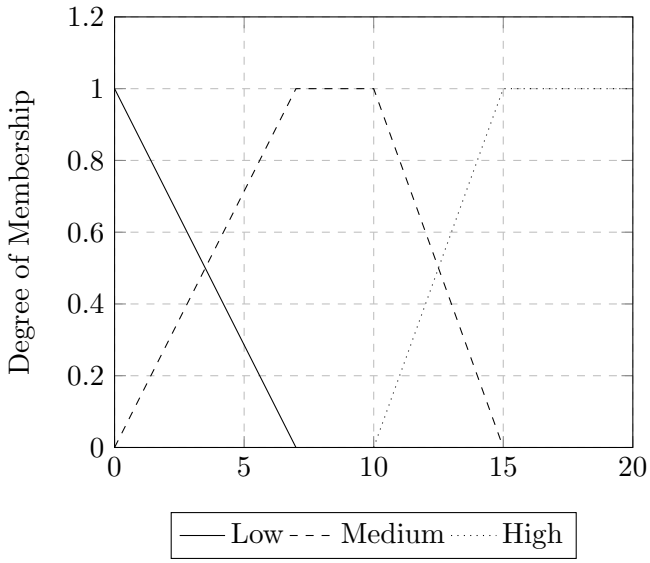

Figure 1: Membership Function Phish Hints

Domain Age

- **Description:** Age of the page in months
- **Integer**. Number 83 in the paper

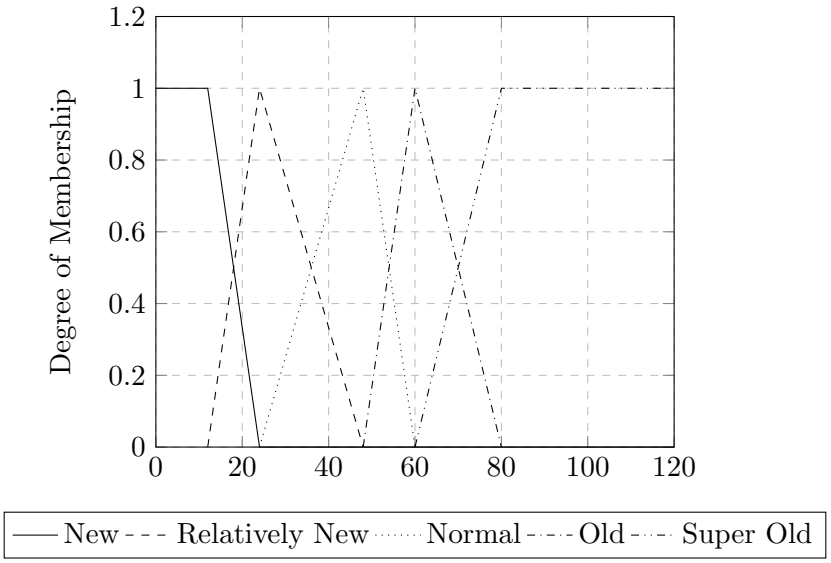

Figure 2: Membership Function Domain Age (in months)

2.1.2 Content features:

Ratio External Hyperlinks

- **Description:** The number of external hyperlinks in a web page divided by the total number of hyperlinks
- **Float**. Number 59 in the paper

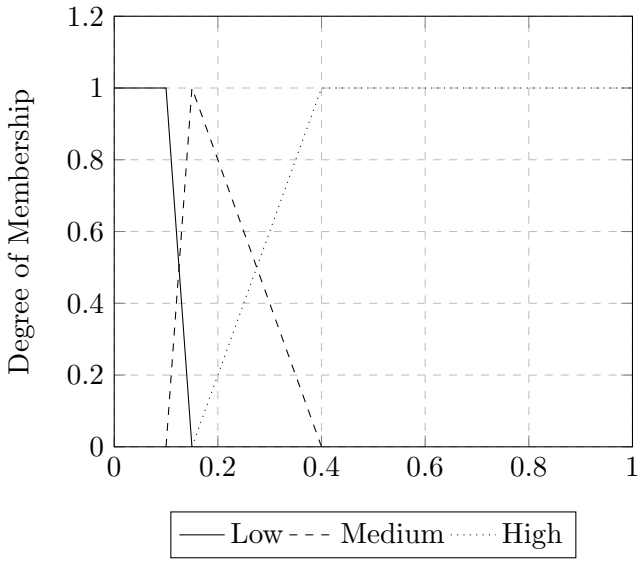

Figure 3: Membership Function Ratio External

2.1.3 External features:

Google Index

- **Description:** Whether a page is indexed in Google
- **Boolean**. Number 86 in the paper

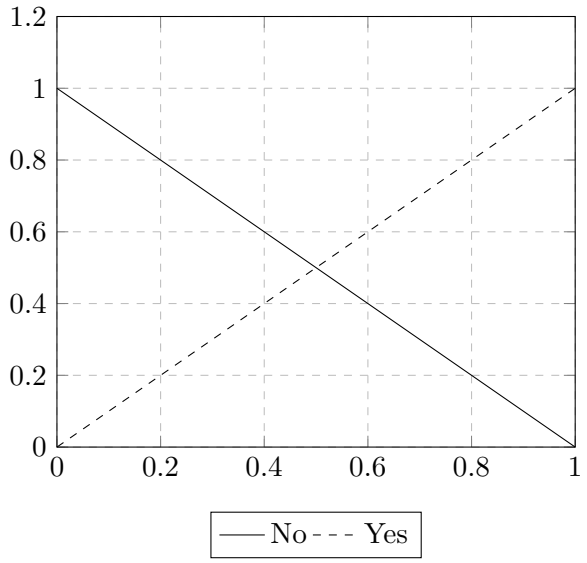

Figure 4: Membership Function Google Index

Page Rank

- **Description:** In the end, we chose Google PageRank which follows the same mantra
- **Integer**. Number 87 in the paper

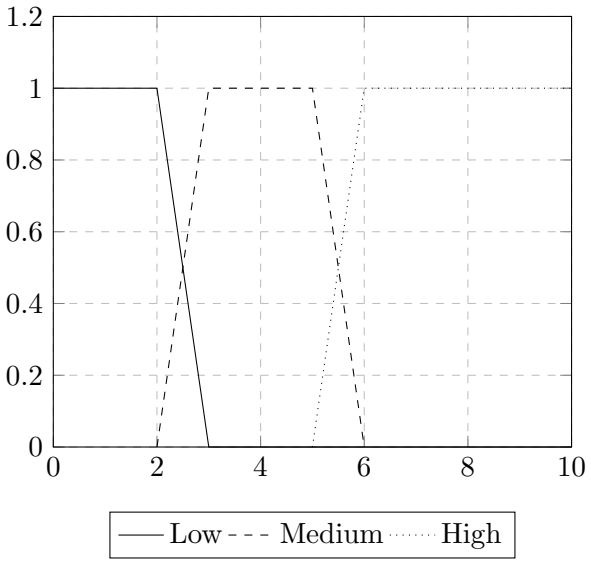

Figure 5: Membership Function Google Page Rank

3 Output Variable

Our output variable will be the phishing risk, where we will consider 5 different fuzzy sets, see 6.

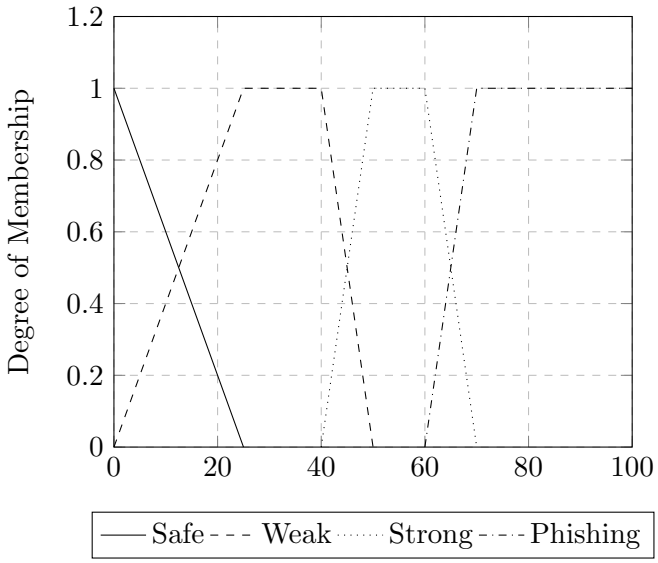

Figure 6: Membership Function Phishing Risk (Output Variable)

4 Rules

5 Implementation

- Mamdami system (min as t-norm and max as t-conorm) - Defuzzification Method: Center of Area - Validate the system using the 3D plot of the rules. ??????????? WTF

6 Testing

- 4 Test case that represent different situation (some must activate more than one lable) - Report the results of each testing case with screenshots and explanations that justify the output obtained (i.e. showing the activations of rules)

7 Complex Fuzzy Expert System

Design (just graphically, no implementation) a more complete fuzzy expert system that includes more features about the websites. Show in a figure the inputs, outputs, and rule blocks that you propose for such expert system. No specific definition of variables nor rules is required.

- We can use a hirarchical rule system

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.