Inteligência Artificial
EIC0029

Ano Letivo 2016/2017

E1: Redes Neuronais para a deteção de sítios web de *phishing*

Relatório Intercalar

Bruno Monteiro Marques
up201405781@fe.up.pt

Vitor Miguel Saraiva Esteves
up201303104@fe.up.pt

02 de Abril de 2017

Objetivo

A prática de *Phishing*[1] consiste na tentativa de obtenção de informação sensível, como por exemplo, *passwords*, *usernames*, números de cartões de crédito, etc, e o seu crescimento exponencial é um problema que pode resultar em roubo de identidade dos utilizadores e/ou danos financeiros, o que leva muitas vezes a perda de confiança dos consumidores em sistemas tecnológicos como e-commerce e online banking.

No final da implementação, o projeto desenvolvido deverá ser capaz de detetar *websites* de *phishing*, através da comparação com um conjunto de exemplos pré-compilados designado de *dataset*, que contém informação acerca das *features* mais comuns que caracterizam estes websites.

Desta forma o objetivo deste trabalho baseia-se na aplicação de Redes Neuronais artificiais para a construção de um método de deteção de sítios web de *phishing*.

[1] https://en.wikipedia.org/wiki/Phishing

Descrição

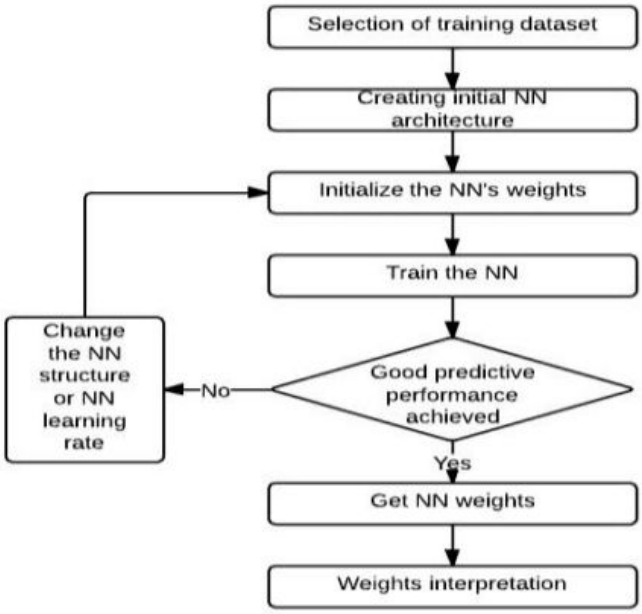
Especificação

Um dataset é um conjunto de dados onde cada elemento pode representar várias características. O *dataset* utilizado consiste numa compilação de 2456 websites, alguns legítimos (colecionados do diretório do Google), outros classificados como *phishers* (colecionados a partir dos arquivos do Phishtank[2] e MillersMiles[3]), das quais foram seleccionados 30 diferentes atributos[4] que podem servir como indicador na deteção de um website de *phishing.* Devido às limitações de extensão deste relatório não iremos apresentar todos os atributos, estes estão descritos na hiperligação indicada acima, no entanto referimos ainda que estes atributos estão divididos entre quatro categorias principais, "*Address Bar based Features*" técnicas apoiadas na análise da utilização dos endereço URL, "*Abnormal Based Features*" técnicas apoiadas na análise de *comportamentos* fora do comum, *"HTML and JavaScript based Features"* técnicas baseadas na exploração de elementos relacionados com o *HTML* e *JavaScript* e por fim "*Domain based Features*" técnicas baseadas na análise do domínio.

A informação agrupada no *dataset* utiliza valores que representam, numericamente usando inteiros, se um determinado atributo se enquadra na categoria de "Legítimo" (1), "Suspeito"(0) ou "*Phishy*"(-1).

[2] https://www.phishtank.com/phish_archive.php
[3] http://www.millersmiles.co.uk/
[4] http://archive.ics.uci.edu/ml/machine-learning-databases/00327/

A estrutura de uma rede neuronal aplicada ao método de detecção de phishing poderia ser a seguinte:



O programa do grupo deve assim treinar apropriadamente uma Rede Neuronal Artificial, usando o algoritmo "Back-Propagation", tendo por base um conjunto de dados disponibilizado para o efeito. Este algoritmo, é o mais usado no treino de ANN's (Artificial Neural Networks). O modelo obtido deve poder depois ser utilizado no diagnóstico de novos casos.

O grupo planeou assim dividir o trabalho da seguinte maneira:

1. Implementar estruturas de apoio que permitam extrair a informação do *dataset* e guardá-la em estruturas apropriadas.
2. Conceber uma rede neuronal multi-camada: a camada de entrada contém os atributos ou variáveis de identificação dos dados, a camada de saída contém a classificação obtida e a(s) camada(s) intermédia(s) auxilia(m) no funcionamento da rede neuronal.
3. Aplicar o algoritmo de *Back-Propagation* aos dados obtidos.
4. Implementar um sistema capaz de uma medição detalhada de resultados nos dados de treino e de teste.

De notar ainda que este trabalho está a ser implementado na linguagem Java, devido à possibilidade de utilização de frameworks e versatilidade da linguagem.

Trabalho efectuado

Inicialmente, o grupo implementou as funcionalidades relativas ao carregamento dos datasets de training e dados. Os dados são guardados em estruturas adequadas para a sua posterior consulta e processamento.

Foi ainda estudado o conteúdo dos *datasets* e a maneira como estavam organizados, de modo a que o grupo percebesse a maneira mais simples de tratar a informação.

Analisou-se ainda a possibilidade de utilizar frameworks externas, como a neuroph[5] e a FANN[6] para facilitar a obtenção de um produto final de melhor qualidade, bem como o funcionamento do algoritmo de *Back-Propagation* e respetiva implementação.

Resultados esperados e forma de avaliação

Os resultados esperados não são fáceis de prever visto que no próprio *dataset é* referida a dificuldade em encontrar training datasets confiáveis, o que prejudicará a forma de avaliação da implementação e a previsão dos resultados. Esperamos ainda poder utilizar outras amostras para avaliar mais corretamente a implementação da Rede Neuronal.

Desta forma espera-se no entanto obter uma rede neuronal capaz de classificar um conjunto de sites como sites legítimos ou sites de phishing.

[5] https://paginas.fe.up.pt/~eol/IA/dokuwiki/doku.php?id=neuroph
[6] https://paginas.fe.up.pt/~eol/IA/dokuwiki/doku.php?id=fann

Conclusões

Após a análise detalhada da informação do trabalho e de um estudo prévio do funcionamento de uma rede neuronal e das suas variadas aplicações, o grupo mostrou-se satisfeito com a opção escolhida. De facto, o projeto é extremamente interessante e útil, não só no âmbito de aprendizagem da cadeira, mas também para outras possíveis iniciativas de índole pessoal.

O tema em concreto, *Phishing,* é também um problema atual que está a aumentar drasticamente como o desenvolvimento das tecnologias modernas e a globalização dos sistemas distribuídos, razão pela qual o grupo ficou contente por poder aprender e possivelmente ajudar na criação de novos métodos de prevenção deste esquema fraudulento.

Finalizando, é ainda de salientar a possível utilização que o grupo irá fazer de *frameworks* de apoio anteriormente desconhecidas, como o *neuroph,* que irão facilitar a implementação do produto desejado.

Recursos

Especificação do trabalho
https://paginas.fe.up.pt/~eol/IA/1617/trabalho.html#E1

Google Docs
https://docs.google.com/

Github
https://github.com

UCI Machine Learning Repository
http://archive.ics.uci.edu/ml/datasets/Phishing+Websites#

Documentação e respetivo *dataset*
http://archive.ics.uci.edu/ml/machine-learning-databases/00327/

Artigo relativo a previsão de *phishing websites* usando Redes Neuronais
https://www.researchgate.net/publication/261636374_Predicting_Phishing_Websites_using_Neural_Network_trained_with_Back-Propagation

Algoritmo de *Back-Propagation*
http://neuralnetworksanddeeplearning.com/chap2.html