Aprendizado de Maquina para Detecção de *Spam*

Thiago Giroto Milani,
Universidade Estadual Paulista - "Julho de Mesquita Filho"
Rio Claro – SP
thiago.milani@unesp.br

Fabricio Aparecido Breve,
Universidade Estadual Paulista - "Julho de Mesquita Filho"
Rio Claro - SP
fabricio@rc.unesp.br

Resumo— Com o grande crescimento da área de informática e inovação tecnológica (era digital) cresce cada vez mais a necessidade de dispositivos e algoritmos capazes de aprender e reconhecer padrões. A segurança computacional se torna cada vez mais essencial com toda essa evolução, pois assim como a tecnologia cresce, os incidentes de segurança crescem duas vezes mais rápido. Com isso surge a necessidade de integrar essas duas vertentes da evolução, aprendizado de maquina e segurança de computadores, trazendo dispositivos e ferramentas capazes de reconhecer padrões de incidentes de segurança através da inteligência computacional.

Área: "Inteligência Computacional".

I. INTRODUÇÃO

Cotidianamente vemos um grande avanço na utilização de algoritmos e modelos computacionais para resolver problemas envolvendo tecnologia. Isso influenciou o crescimento de pesquisas na área de aprendizado de máquina e inteligência artificial, sendo cada vez mais frequente a criação de grupos e eventos internacionais e nacionais ligados ao assunto.

Cada vez mais os usuários de computador têm-se deparado com problemas mais sérios de segurança de computadores e redes. Com o crescimento da internet desde sua criação, e cada vez mais serviços e aplicações para ela, acabou sendo um bem essencial e de extrema necessidade, trazendo assim mais destaque para pessoas mal-intencionadas disseminarem vírus, *malware* e *spans* com intenção de roubar ou atacar algum usuário ou empresa. São cada vez mais comuns notícias de invasão de sites.

Devido ao tamanho e quantidade de informação redundante das mensagens enviadas por *e-mail* se dá a necessidade de fazer a extração das características mais relevantes, sendo um dos pontos principais da pesquisa pois devido a essa etapa o algoritmo de classificação poderá ser executado mais rapidamente para classificando a mensagem como *spam* ou legítima.

A técnica de extração de características se baseia em transformar dados de entrada muito grande (muitos dados, porém pouca informação) para serem processados em conjuntos reduzidos de características melhores representadas (também chamados de vetores). Se essas características tiverem sido extraídas cuidadosamente se espera que esse conjunto represente a parte mais relevante da informação, tendo assim uma classificação mais confiável.

Além de já trabalhar com segurança computacional a algum tempo, com experimentos utilizando *honeypot* [7] uma motivasão pessoal para este artigo é a grande quantidade de ataques e mensagens de spam que são disseminados todos os dias, além dos recentes acontecimentos em escala global como o *malware wannacry*,[8] o qual ocasionou a paralisação de diversos serviços essenciais totalmente dependentes da tecnologia e internet.

O objetivo da pesquisa é analisar algoritmos de extração de características combinados com algoritmos clássicos de classificação utilizando o aprendizado de máquina, para tenta uma melhor classificação de mensagens de *spam*.

II. CONCEITOS E TÉCNICAS

A.Extração de Características

A extração de características engloba simplificar o conjunto de dados requeridos para descrever um grande conjunto com mais precisão

Extração de características é um termo genérico para os métodos de construção e combinações de valores, transformando uma entrara de dados grande, em um vetor simplificado, onde seu processamento é mais rápido.

Os melhores resultados são obtidos na extração de características quando um especialista gera um conjunto de características conforme a aplicação, porém quando não se tem essa possibilidade é utilizado técnicas de redução e redimensionamento genéricas, tais como:

- *Principal Components Analysis;*
- *Semidefinite Embedding;*
- *Multifactor Dimensionality Reduction;*
- *Nonlinear Dimensionality Reduction;*
- *Isomap;*
- *Kernel PCA;*
- *Latent Semantic Analysis;*
- *Partial Least Squares;*
- *Independent Component Analysis.*

B.*Mensagens de Spam*

Uma grande maioria dos ataques computacionais, roubos de informação, e invasões de sistemas e redes se originam de mensagens de *spam* não detectadas ou pouco conhecimento por parte dos usuários para detecção manual desse tipo de mensagem.

Uma mensagem de *spam* pode ser de vários tipos como um *e-mail* falso, com informações genéricas com a intenção de enganar o leitor a ponto de extrair informações sigilosas, envio de propagandas não autorizadas, ou até mesmo ocasionar a instalação de aplicativos e ferramentas que irão proporcionar o roubo de informação ou acesso indevido a máquina e arquivos sigilosos (Figura 1).

Uma definição mais formal diz que *spam* é o termo usado

VIII Workshop do Programa de Pós-Graduação em Ciência da Computação: "Interação entre Academia e Empresa", Unesp, Presidente Prudente, 12 e 13 de novembro de 2018.

para se referir aos *e-mails* não solicitados, que geralmente são enviados para um grande número de pessoas. Quando este tipo de mensagem possui conteúdo exclusivamente comercial também é referenciado como UCE (*Unsolicited Commercial E-mail*). [1].

A técnica de mensagens de *spam* utiliza de engenharia social para enganar o usuário, podendo agir de diversas formas diferentes quando executado, conforme mencionado em [1] ocasionando:

- Perda de mensagens importantes;
- Conteúdo impróprio ou ofensivo;
- Não recebimento de *e-mails*;
- Classificação errada de mensagens.

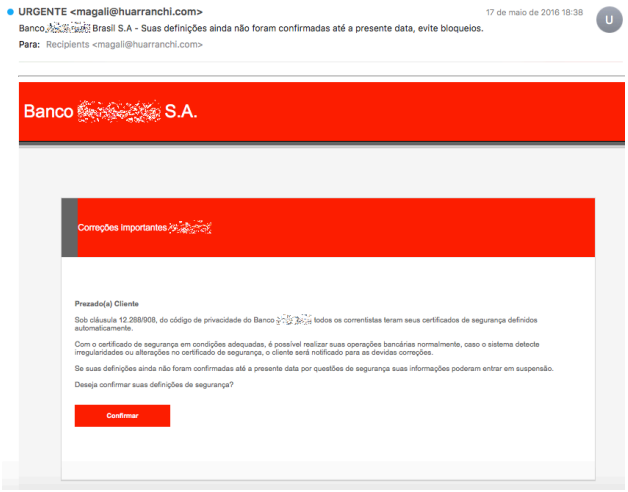

Figura 1 - Captura de tela de um e-mail com uma mensagem de spam.

C.Aprendizado de Máquina

A utilização de máquinas para realizar diversas tarefas quotidianas ainda é algo muito recente. No entanto, nos últimos tempos isso vem se tornando mais comum. Com o aprendizado de máquina na área de inteligência artificial, as máquinas vêm conseguindo executar várias funções que, até então, apenas humanos eram capazes. A principal ideia do aprendizado de máquina é a criação de mecanismos que sejam capazes de imitar não apenas a mente humana [2] mas a de diversos outros seres vivos. Essa técnica abrange explorar os mais diversos campos, presente em praticamente todas as áreas de estudo. Na maioria das aplicações, o foco principal é o reconhecer padrões, onde os seres humanos são muito bons. Uma das vantagens na utilização de métodos computacionais para o reconhecimento de padrões está na precisão que pode ser obtida com essa tarefa, além da capacidade de processar uma grande quantidade de informação em um curto período de tempo.

O aprendizado de máquina é o desenvolvimento de algoritmos que podem aprender e se adaptar. Diferentemente dos programas convencionais, isto nos traz uma capacidade de que o algoritmo melhore seu desempenho através do aprendizado de experiências anteriores [3], o qual se dá com o maior número de dados possível. Com isso e a evolução dos métodos de aprendizado de máquina, os dados disponibilizados na Internet para testes aumentaram consideravelmente [4].

Os métodos de aprendizado de máquina para classificação podem ser divididos em duas principais categorias: aprendizado supervisionado e aprendizado não-supervisionado.

III. METODOLOGIA DE DESENVOLVIMENTO

A classificação de mensagens de *spam* é algo que até hoje não é bem feito pelas ferramentas disponíveis, justamente pela grande variedade de tipos e técnicas de mensagens de *spam*. Para se classificar uma mensagen como *spam* ou legítima normalmente é utilizado de uma base de dados com palavras chave as quais uma determinada combinação pode ou não a classificar como *spam*.

A metodologia que será abordada neste trabalho é a comparação da combinação de extratores de características com algoritmos de classificação. Ou seja, será feito a extração de características de um *dataset* com algoritmos distintos, e posteriormente classificados por diferentes algoritmos clássicos de classificação, utilizando aprendizado de maquina a fim de achar a melhor combinação possível entre extrator de características e classificador, com a finalidade de detectar o máximo de mensagens de *spam*.

IV. CONSIDERAÇÕES FINAIS

Com este projeto será possível apresentar tanto uma visão geral quanto específica, sobre os temas que englobam este trabalho. A busca por uma maneira mais confiável de classificar as mensagens de *spam* tem uma grande necessidade devido a velocidade de evolução das tecnologias e ataques utilizando essas técnicas. Com isso será possível apresentar a melhor combinação entre os principais extratores de características e os principais algoritmos de classificação utilizando o aprendizado de máquina para a finalidade de detecção e classificação de mensagens de *spam*.

REFERÊNCIAS

[1] Cert.br. - Centro de estudos, resposta e tratamento de incidentes de segurança no brasil, cartilha de segurança para internet. Cartilha de Segurança para a Internet, 2016, 2016. URL https://cartilha.cert.br;

[2] J. de Fernandes Teixeira - Inteligência artificial. Pia Sociedade de São Paulo-Editora Paulus, 2014;

[3] A. Blum. - Machine learning theory. Carnegie Melon Universit, School of Computer Science, page 26, 2007;

[4] S. Dumais, J. Platt, D. Heckerman, and M. Sahami. - Inductive learning algorithms and representations for text categorization. In Proceedings of the seventh international conference on Information and knowledge management, pages 148–155. ACM, 1998;

[5] M. N. Murty, A. Jain, and P. Flynn. - Data clustering: a review acm compt. surv. ACM Computing Surveys, 31(3), 1999;

[6] V. Moll. - Detecção de intrusão uzando técnica de aprendizado de m aquinas, 2010. Dissertação Mestrado Universidade Federal de Santa Catarina, Florianópolis – SC;

[7] T. G. Milani. - Honeypot para detecção e contenção de invasões e Botnets, 2011. Monografia (Bacharel em Sistemas de Informação), UNICEP (Centro Universitário Central Paulista), São Carlos, Brasil;

[8] M. Couto. - Ataque hacker global ´e sinal de alerta para empresas e governos. https://exame.abril.com.br/revista-exame/um-sinal-de-alerta/ 2017. Revista EXAME;

[9] J. de Fernandes Teixeira - Inteligência artificial. Pia Sociedade de São Paulo-Editora Paulus, 2014.