

Identificação de alertas de segurança virtual veiculados no Twitter

Jackson J. de Souza

Orientador: Daniel M. Batista Coorientadora: Elisabeti Kira

Instituto de Matemática e Estatística

IME-USP

27 de Janeiro de 2015



IME - Instituto de
Matemática e Estatística

Introdução

- Computadores e internet bastante popularizados
- Redes sociais online também
- Nova gama de crimes possíveis
- Formas atuais de combate e prevenção são insuficientes

- Artigo de Rodrigo Campiolo e Luiz Arthur, ambos da UTFPR
- Foi confirmada a disseminação rápida e confiável de alertas de segurança virtual (ASVs) no Twitter
- ASVs aparecem no twitter antes de alguns sites de mídia especializada

Proposta

- Criação de um modelo de aprendizagem computacional que identifica ASVs no Twitter em língua inglesa utilizando a Weka
- Para isso será feita uma comparação de desempenho entre os classificadores *SVM* e *Naive Bayes*



Objetivo

- Identificar o surgimento de alertas de segurança virtual (ASVs) por meio de postagem de mensagens
- Utilização do Twitter como fonte de informação
- Aplicação de aprendizagem de máquina para identificar ASVs

Aprendizagem de máquina

- O que é?
- Como funciona?
- O que caracteriza um bom aprendiz?

Exemplo de aprendizagem



Características

- Tamanho
- Cor
- Formato
- etc.

- **Notícia sobre segurança virtual**

“Antivirus News: McAfee mocks McAfee antivirus in video rant
WA today: McAfee mocks McAfee antivirus in vid...
<http://t.co/nuhVDMpU48>”

- **Alerta de segurança virtual**

“Malware spread on Skype taps victim PCs to mint #bitcoin
<http://t.co/snc8yqGIL9> by @ArsTechnica's @dangoodin001”

Classes do problema

- **Notícia sobre segurança não-virtual**

“RT @Timodc: RT @Timodc: As Biden makes a absurdly false attack on Mitt, guess who has supported social security taxes? Biden. <http://t.co/5YCI3etf>”

- **Spam**

“Choose awardwinning security for your #PC. Choose @McAfee to protect against viruses, malware, and other threats. <http://t.co/xCDrIMkl>”

Coleta e preparação dos dados

- Foram coletados 9403 tuítes entre outubro de 2013 e junho de 2014
- Tuítes foram separados em 4 classes mutuamente excludentes
- Geração de arquivo .arff com tuítes processados
 - Decodificação dos tuítes
 - Remoção de metadados, exceto hashtags
 - Busca de alguns padrões identificados nos tuítes

Experimentos

- Leitura dos tuítes pela Weka e aplicação de um filtro que divide os tuítes em tokens
- Information Gain
- Conjunto de tuítes dividido em 66% (conjunto de treinamento) e 33% (conjunto de testes)

Naive Bayes

Tuítes corretamente classificados	2507	79.16%
Tuítes incorretamente classificados	660	20.84%

Taxa VP	Taxa FP	Precision	Recall	F-Measure	Classe
0.456	0.054	0.569	0.456	0.507	Notícia de segurança virtual
0.858	0.169	0.816	0.858	0.837	Alerta de segurança virtual
0.883	0.059	0.887	0.883	0.885	Notícia de segurança geral
0.474	0.035	0.439	0.474	0.456	Spam
0.792	0.108	0.787	0.792	0.788	Média ponderada das classes

Support Vector Machines (SVM)

Tuítes corretamente classificados	2542	80.26%
Tuítes incorretamente classificados	625	19.73%

Taxa VP	Taxa FP	Precision	Recall	F-Measure	Classe
0.374	0.034	0.628	0.374	0.469	Notícia de segurança virtual
0.907	0.202	0.797	0.907	0.848	Alerta de segurança virtual
0.897	0.064	0.881	0.897	0.889	Notícia de segurança geral
0.37	0.019	0.525	0.37	0.434	Spam
0.803	0.122	0.789	0.803	0.789	Média ponderada das classes

Proposta de 2 modelos de classificação de tuítes que possuem uma alta taxa de recuperação de ASVs

Trabalhos futuros

- Utilizar o Open Calais para identificar entidades nos tokens dos tuítes para melhorar o classificador
- Desenvolver um software que classifica tuítes logo após a postagem deles
- Estudo de mais classificadores para identificar ASVs

Referências I



Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten.

The WEKA data mining software: an update.

SIGKDD Explorations, 11, 2009.

Issue 1.



Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze.

An Introduction to Information Retrieval.

Cambridge University Press, draft edition, April 2009.



LUIZ ARTHUR F. SANTOS, Rodrigo CAMPIOLO,
MARCO AURELIO GEROSA, and DANIEL MACEDO
BATISTA.

Análise de mensagens de segurança postadas no twitter.
*Anais do simpósio brasileiro de sistemas colaborativos
(SBSC), (3):20–28, 2012.*