Universidade Federal de Pernambuco
Centro de Informática
Graduação em Ciência da Computação

Detecção de Posicionamento no contexto de *Fake News*

Larissa Navarro Passos de Araujo
Proposta de Trabalho de Graduação

Orientador: Tsang Ing Ren

Recife
Março de 2016

Resumo

A última campanha presidencial americana trouxe à tona o quão grave o problema de *fake news* havia se tornado. A desconfiança da população em relação às grandes organizações de notícias aliadas a redes sociais transformadas em câmaras de eco formou um ambiente propício para a disseminação de notícias falsas, sem a devida verificação. A detecção de *fake news* é uma tarefa extremamente complicada, mesmo para humanos, e a automatização de tal problema ainda está distante. Um primeiro passo nesse sentido é a automatização da detecção de posicionamento de textos em relação a outros, a base do *Fake News Challenge*, a primeira competição de Processamento de Linguagem Natural na área. O objetivo deste trabalho é implementar um sistema de classificação de posicionamento de um corpo de texto em relação a uma manchete. O texto pode concordar, discordar, discutir ou não ser relacionado à manchete. Para isso, serão utilizadas técnicas clássicas de Aprendizagem de Máquina e Processamento de Linguagem Natural.

Palavras-chave: processamento de linguagem natural, aprendizagem de máquina, engenharia de características.

Abstract

The last American presidential run has brought the Fake News problem to the light. The growing weariness against the great media conglomerates, along with the eco chambers of social networks, has developed an ideal environment for the spreading of Fake News. Fake News detection is an extremely difficult task, even for humans. As such, automation is still a long way away. Stance Detection automation is a first step in that Direction. It is the basis of the Fake News Challenge, the first Natural Language Processing competition in this field. The goal of this work is to implement a stance detection system, capable of classifying the stance of a body of text relative to a headline. The text can agree, disagree, discuss or be unrelated to the headline. Classical Machine Learning and Natural Language Processing techniques will be utilized.

Keywords: machine learning, natural language processing, feature engineering.

Sumário

1. Introdução 1
2. Objetivos 2
3. Cronograma 3
4. Possíveis Avaliadores 4
5. Referências 5
6. Assinaturas 6

1. Introdução

Desde a última campanha presidencial americana o termo *fake news* tornou-se um grande problema tanto para a indústria de notícias, como para grandes redes sociais como *Facebook* e *Twitter*. Segundo o *New York Times*, o termo é aplicável a uma matéria inventada com a intenção de ludibriar o público [1]. Uma pesquisa feita em dezembro de 2016 pelo instituto *Pew Research* revelou que 64% dos adultos norte-americanos disseram ter se sentido muito confusos sobre eventos correntes, devido a notícias inventadas [2].

Tanta incerteza quanto à veracidade de informações levou ao crescimento de organizações de *fact checking*, ou seja, a tarefa de determinar a veracidade e corretude de afirmações textuais. Esta tarefa é extremamente complexa, e mesmo especialistas enfrentam dificuldades ao executar esta tarefa. Assim, seria de grande valia para diversos setores uma versão automatizada de *fact checking*. No entanto, esta tarefa ainda está muito longe de ser resolvida no campo do Processamento de Linguagem Natural. Desafios existentes incluem a necessidade de aplicação de contexto, adquirir e verificar fontes de verdade, falta de dados de treinamento, entre outros [3].

Neste contexto, surgiu o *Fake News Challenge* [4], uma competição internacional na área de Inteligência Artificial, que busca resolver a primeira parte do problema de *fact checking*: definir o posicionamento de dois textos em relação a um assunto, afirmação ou problema, a chamada detecção de posicionamento. Na competição, o problema será reduzido à tarefa de detectar o posicionamento de um corpo de texto de notícia relativo à uma manchete. O corpo de texto poderá concordar, discordar, discutir ou não ser relacionado à manchete.

Uma boa solução para o problema de detecção de posicionamento permitiria a um *fact Checker* humano submeter uma manchete e recuperar todos os textos com posicionamentos relativos a esta manchete, facilitando seu trabalho de recuperar fontes relevantes. Pode-se usar detecção de posicionamento também para classificar uma matéria como falsa ou verdadeira, utilizando a credibilidade ponderada aplicada ao posicionamento de outras fontes sobre o assunto.

2. Objetivos

O objetivo deste trabalho é construir um sistema que possa classificar o posicionamento relativo entre um corpo de texto e uma manchete, utilizando técnicas clássicas de Aprendizagem de Máquina (como máquinas de vetores suporte – *SVM*) e Processamento de Linguagem Natural, afim de participar do *Fake News Challenge*.

3. Cronograma

As atividades pretendem ser desenvolvidas de acordo com o cronograma abaixo:

Atividade	Março				Abril				Maio				Junho				Julho			
Revisão Bibliográfica																				
Implementação do classificador																				
Análise de resultados																				
Elaboração do relatório final																				
Preparação da defesa																				
Defesa																				

4. Possíveis Avaliadores

São possíveis avaliadores do trabalho os professores abaixo:

- Prof. George Darmiton;
- Prof. Paulo Salgado;
- Prof. Luciano Barbosa.

5. Referências

[1] As Fake News spreads lies, more readers shrug at the Truth. Acesso em: 21/03/2017. Disponível em: https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html

[2] Many Americans believe Fake News is sowing confusion. Acesso em: 21/03/2017. Disponível em: http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

[3] Introducing Factmata – Artificial Intelligence for automated fact-checking. Acesso em: 26/03/2017. Disponível em: https://medium.com/factmata/introducing-factmata-artificial-intelligence-for-political-fact-checking-db8acdbf4cf1

[4] Fake News Challenge. Acesso em: 26/03/2017. Disponível em: http://www.fakenewschallenge.org/

6. Assinaturas

Tsang Ing Ren
Orientador

Larissa Navarro Passos de Araujo
Aluna