



Universidade Federal de Pernambuco
Centro de Informática

Graduação em Ciência da Computação

***Deep Learning* na Detecção de
Posicionamento em Notícias *Online***

João Pedro de Carvalho Magalhães

Proposta de Trabalho de Graduação

Orientador: Prof. Tsang Ing Ren

Recife
Abril de 2017

Resumo

O compartilhamento de notícias falsas vêm aumentando demasiadamente na *web*, principalmente em redes sociais. Isso é alarmante pois informações falsas repassadas e entendidas como verdadeiras causam um forte impacto negativo na sociedade. Portanto, existe a necessidade de se criar um sistema automatizado capaz de detectar afirmações falsas para que se evite a acelerada propagação desse tipo de notícia. Este trabalho propõe que uma das partes desse sistema seja um classificador de abordagem *deep learning* que detecte posicionamento opinativo em notícias *online*. Objetiva-se que tal classificador seja implementado, sua performance analisada, e seu papel no sistema discutido.

Palavras-chave: notícias online, notícias falsas, detecção de posicionamento, classificação de texto, deep learning, processamento de linguagem natural, aprendizagem de máquina

Abstract

Fake news sharing has overly increased in the web, mainly in social networks. This is threatening because false information which are passed on and believed to be true cause a strong negative impact to society. Thus, there is a need to create an automated system capable of detecting false claims in order to avoid an accelerated propagation of this kind of news. This work proposes one part of this system to be a classifier based on deep learning for stance detection in online news. The goal is to implement such classifier, analyze its performance, and discuss its role in the system.

Keywords: online news, fake news, stance detection, text classification, deep learning, natural language processing, machine learning

Sumário

1	Contexto	1
2	Objetivo	2
3	Cronograma	3
4	Possíveis Avaliadores	4
5	Assinaturas	5

CAPÍTULO 1

Contexto

A *World Wide Web* é uma enorme fonte de dados. A cada segundo que passa milhares de novos arquivos e informações são colocadas *online* por pessoas do mundo inteiro. E essa produção vêm crescendo ainda mais com a popularização de *blogs* e redes sociais, pois agora cada indivíduo possui um espaço próprio e interconectado para elaborar e compartilhar seu conteúdo ou o de outras pessoas e veículos de comunicação. E ele pode fazer isso de qualquer lugar se em posse de um *smartphone* e acesso à internet móvel. Ou seja, qualquer pessoa nessas condições é um divulgador instantâneo e global de informações, sejam elas factuais ou opinativas.

Esse poder que cada um tem, apesar de impressionante e animador, também pode ser perigoso quando os usuários não conseguem checar a veracidade das informações [1], o que é complicado até para alguns experts [2]. Essa incapacidade junto da acelerada propagação de notícias ou informações falsas podem afetar a sociedade negativamente, influenciando a opinião pública e até o mercado financeiro [3][4][5].

Neste conjuntura, grandes empresas estão se juntando para estudar o problema [6], e a detecção de posicionamento surgiu como um passo importante para possibilitar o avanço na detecção automática de notícias falsas [7]. A detecção de posicionamento é a tarefa de identificar o tipo de relação entre dois discursos. Um discurso pode ser apenas um tópico, um argumento ou uma afirmação, e essa relação pode ser de concordância, discordância, ou neutralidade – quando o discurso de um autor mostra que este não têm uma posição opinativa acerca do discurso do outro. Utilizando essa relação, pode-se comparar o que veículos diferentes estão dizendo sobre um mesmo tópico em suas notícias.

A automação da tarefa pode ser alcançada através de abordagens computacionais. Uma que vêm tendo sucesso em tarefas de detecção, passando a ser bastante usada em processamento de linguagem natural, é chamada de *deep learning*. Essa técnica consiste em adicionar muitas camadas escondidas em uma rede neural para aumentar o aprendizado sobre a tarefa, exigindo um grande poder de processamento e vasta quantidade de dados.

CAPÍTULO 2

Objetivo

O propósito deste trabalho é construir um classificador que utilize uma abordagem *deep learning* para determinar o tipo de posicionamento opinativo que títulos de notícias online têm com relação às afirmações presentes em seu corpo. Ele também responderá se o posicionamento está de fato presente, e por esse motivo, essa tarefa descrita inicialmente em [7] estende a proposta feita por [8].

Pretende-se também realizar experimentos sobre as *features* e técnicas de *deep learning* para poder analisar a performance do classificador implementado com respeito ao *baseline* existente [9]. O conjunto de dados a ser utilizado contém somente notícias na língua inglesa e em forma de texto [10].

Além disso, propõe-se discutir o papel desse classificador no processo de identificação automatizada de notícias falsas *online*.

CAPÍTULO 3

Cronograma

Atividades	Março	Abril	Maio	Junho	Julho
Formulação da Proposta	█	█			
Revisão Bibliográfica	█	█			
Implementação do Classificador		█	█		
Testes e Experimentos			█	█	
Elaboração do Relatório Final				█	█
Preparação da Defesa					█
Defesa					█

Cronograma baseado no calendário acadêmico disponível no site da UFPE [11].

CAPÍTULO 4

Possíveis Avaliadores

Um possível avaliador do trabalho proposto é o professor Luciano de Andrade Barbosa (CIn/UFPE).

CAPÍTULO 5

Assinaturas

João Pedro de Carvalho Magalhães
(Aluno)

Tsang Ing Ren
(Orientador)

Referências Bibliográficas

- [1] Many Americans Believe Fake News Is Sowing Confusion. Disponível em: <<http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>>. Acesso em: 4 abr. 2017.
- [2] Introducing Factmata — Artificial intelligence for automated fact-checking. Disponível em: <<https://medium.com/factmata/introducing-factmata-artificial-intelligence-for-political-fact-checking-db8acdbf4cf1>>. Acesso em: 4 abr. 2017.
- [3] Fake news: an insidious trend that’s fast becoming a global problem. Disponível em: <<https://www.theguardian.com/media/2016/dec/02/fake-news-facebook-us-election-around-the-world>>. Acesso em: 4 abr. 2017.
- [4] Can ‘Fake News’ Impact The Stock Market?. Disponível em: <<https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/#17700ba12fac>>. Acesso em: 4 abr. 2017.
- [5] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” tech. rep., National Bureau of Economic Research, 2017.
- [6] What is fake news, what are Facebook and Google doing about it, why is the Government boycotting YouTube and how can I tell if a site is legitimate?. Disponível em: <<https://www.thesun.co.uk/news/2188911/fake-news-inquiry-facebook-google-boycott/>>. Acesso em: 4 abr. 2017.
- [7] FNC - Fake News Challenge. Disponível em: <<http://www.fakenewschallenge.org/>>. Acesso em: 2 fev. 2017.
- [8] W. Ferreira and A. Vlachos, “Emergent: a novel data-set for stance classification,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ACL, 2016.
- [9] A baseline implementation for FNC-1. Disponível em: <<https://github.com/FakeNewsChallenge/fnc-1-baseline>>. Acesso em: 10 mar. 2017.
- [10] Stance Detection dataset for FNC-1. Disponível em: <<https://github.com/FakeNewsChallenge/fnc-1>>. Acesso em: 7 fev. 2017.

- [11] Calendário Acadêmico 2017 - UFPE. Disponível em: <https://www.ufpe.br/proacad/images/Calendario_Academico/UNIVERSIDADE_FEDERAL_DE_PERNAMBUCO__calend%C3%A1rio_acad%C3%AAmico_2017.pdf>. Acesso em: 4 abr. 2017.
- [12] Nas redes, mentiras sobre pleito nos EUA superam notícias reais. Disponível em: <<http://ww1.folha.uol.com.br/mundo/2016/11/1833158-nas-redes-mentiras-sobre-pleito-nos-eua-superam-noticias-reais.shtml/>>. Acesso em: 4 abr. 2017.