XXIV Encontro de Iniciação Científica IX Encontro de Iniciação em Desenvolvimento Tecnológico e Inovação



24 a 26 de setembro

IDENTIFICAÇÃO DE LINGUAGEM OFENSIVA EM MÍDIAS SOCIAIS UTILIZANDO ALGORITMOS DE APRENDIZADO DE MÁQUINA



Murilo de Paula Araujo

(Faculdade de Engenharia de Computação; Bolsista PIBIC/CNPq; murilodepa@gmail.com)
Orientador: Prof. Dr. Juan Manuel Adán Coello

(Grupo de Pesquisa em Sistemas Inteligentes; Linha de Pesquisa: Sistemas Inteligentes Distribuídos; CEATEC. juan@puc-campinas.edu.br)

INTRODUÇÃO

- Agressão e abuso online, como discursos de ódio em mídias sociais, têm trazido muitos inconvenientes aos usuários, levando alguns à desativação de contas, à autoflagelação e mesmo ao suicídio;
- Utilização de métodos de processamento de língua natutal (PLN) e algoritmos de aprendizado de máquina.

METODOLOGIA

- Estudo sobre detecção de posicionamento e identificação de linguagem ofensiva em mídias sociais;
- Obtenção de conjuntos de dados rotulados contendo linguagem ofensiva e não ofensiva originária de mídias sociais;
- Estudo e aplicação de algoritmos de aprendizado de máquina.

PALAVRAS CHAVE

Linguagem ofensiva, discurso de ódio, detecção de posicionamento, aprendizado de máquina, aprendizado profundo, deep learning, mídias sociais.

OBJETIVO

Identificação semiautomática da presença de linguagem ofensiva em textos publicados em mídias sociais utilizando métodos de processamento de língua natural (PLN) e algoritmos de aprendizado de máquina.

RESULTADOS ESPERADOS

- Classificadores para a identificação de linguagem ofensiva em mídias sociais;
- Análise da acurácia dos classificadores;
- Conjuntos de dados para o desenvolvimento de novas pesquisas sobre o tema.



BIBLIOGRAFIA BASICA

- R. Kumar, G. Bhanodai, R. Pamula, M. R. Chennuru, "TRAC-1 Shared Task on Aggression Identification: IIT (ISM) COLING'18", in Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), 2018, p. 58–65.
- Z. Waseem, T. Davidson, D. Warmsley, I. Weber, "Understanding abuse: A typology of abusive language detection subtasks", arXiv preprint arXiv:1705.09899, 2017.
- A. D. Costa Neto, J. M. Adán Coello. Redes Neurais Convolucionais Aplicadas à Análise de Sentimento. In: XXII Encontro de Iniciação Científica da PUC-Campinas, 2017, Campinas.
- P. Grandin e J. M. Adán, "Piegas: A systems for sentiment analysis of tweets in portuguese", IEEE Latin America Transactions, vol. 14, nº 7, p. 3467–3473, 2016.
- P. C. T. Fortuna, "Automatic detection of hate speech in text: an overview of the topic and dataset annotation with hierarchical classes", Dissertação de Mestrado Integrado em Engenharia Informática e Computação, Universidade do Porto, 2017.