



Título: Automatic Misinformation Detection About COVID-19 in Brazilian Portuguese WhatsApp Messages

Data: 02/05/2023

Horário: 16h00

Local: Videoconferência

Resumo:

During the COVID-19 pandemic, the problem of misinformation arose once again, quite

intensely, through social networks. In many developing countries such as Brazil, one of the primary sources of misinformation is the messaging application WhatsApp. However, due to WhatsApp's private messaging nature, there still few methods of misinformation detection developed specifically for this platform. Additionally, a MID model built to Twitter or Facebook

may have a poor performance when used to classify WhatsApp messages. In this context, the automatic misinformation detection (MID) about COVID-19 in Brazilian Portuguese WhatsApp messages becomes a crucial challenge. In this work, we present the COVID- 19.BR, a data set of WhatsApp messages about coronavirus in Brazilian Portuguese, collected from Brazilian public groups and manually labeled. Besides, we evaluated a series of misinformation

classifiers combining different techniques. Our best result using classic machine learning achieved an F1 score of 0.799, and the analysis of errors indicates that they occur mainly due to the predominance of short texts. When texts with less than 50 words are filtered, the F1 score rises to 0.866. Additionally, we propose a new approach, called MIDeepBR, based on BiLSTM neural networks, pooling operations and attention mechanism, which is able to automatically detect misinformation in Brazilian Portuguese WhatsApp messages. MIDeepBR outperforms classical machine learning approaches achieving an F1 score of 0.834. Finally, we explore a post-hoc interpretability method called LIME to explain the predictions of the MID approaches. Besides, we apply a textual analysis tool called LIWC to analyze WhatsApp messages' linguistic characteristics and identify psychological aspects present in misinformation and non-misinformation messages. The results indicate that it is feasible to understand relevant aspects of the MID model's predictions and find patterns on WhatsApp messages about COVID19. So, we hope that these findings help to understand the misinformation phenomena about COVID-19 in WhatsApp messages.

Banca examinadora:

- Prof. Dr. Javam de Castro Machado (MDCC/UFC - Orientador)
- Prof. Dr. José Maria da Silva Monteiro Filho (UFC - Coorientador)
- Prof. Dr. César Lincoln Cavalcante Mattos (UFC)
- Prof. Dr. Victor Aguiar Evangelista de Farias (UFC)