World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:15, No:04, 2021

Improving Subjective Bias Detection Using Bidirectional Encoder Representations from Transformers and Bidirectional Long Short-Term Memory

Authors : Ebipatei Victoria Tunyan, T. A. Cao, Cheol Young Ock

Abstract : Detecting subjectively biased statements is a vital task. This is because this kind of bias, when present in the text or other forms of information dissemination media such as news, social media, scientific texts, and encyclopedias, can weaken trust in the information and stir conflicts amongst consumers. Subjective bias detection is also critical for many Natural Language Processing (NLP) tasks like sentiment analysis, opinion identification, and bias neutralization. Having a system that can adequately detect subjectivity in text will boost research in the above-mentioned areas significantly. It can also come in handy for platforms like Wikipedia, where the use of neutral language is of importance. The goal of this work is to identify the subjectively biased language in text on a sentence level. With machine learning, we can solve complex AI problems, making it a good fit for the problem of subjective bias detection. A key step in this approach is to train a classifier based on BERT (Bidirectional Encoder Representations from Transformers) as upstream model. BERT by itself can be used as a classifier; however, in this study, we use BERT as data preprocessor as well as an embedding generator for a Bi-LSTM (Bidirectional Long Short-Term Memory) network incorporated with attention mechanism. This approach produces a deeper and better classifier. We evaluate the effectiveness of our model using the Wiki Neutrality Corpus (WNC), which was compiled from Wikipedia edits that removed various biased instances from sentences as a benchmark dataset, with which we also compare our model to existing approaches. Experimental analysis indicates an improved performance, as our model achieved state-of-the-art accuracy in detecting subjective bias. This study focuses on the English language, but the model can be fine-tuned to accommodate other languages.

Keywords : subjective bias detection, machine learning, BERT–BiLSTM–Attention, text classification, natural language processing
Conference Title : ICEMNLP 2021 : International Conference on Empirical Methods in Natural Language Processing
Conference Location : Venice, Italy
Conference Dates : April 12-13, 2021