Sarcasm Detection in Arabic Tweets: A comparison Between deep learning and Pre trained Transformers-based Models

Rania Bouguesri
Laboratoire de Méthodes de Conception des Systèmes (LMCS)
Ecole Nationale Superieure d'Informatique (ESI)
Algiers, Algeria
hr_bouguesri@esi.dz

Khadidja Habelhames
Laboratoire de Méthodes de Conception des Systèmes (LMCS)
Ecole Nationale Superieure d'Informatique (ESI)
Algiers, Algeria
hk_hab_el_hames@esi.dz

Hassina Aliane
Research and Development in Digital Humanities Division
Research Center on Scientific and Technical Information (CERIST)
Algiers, Algeria
ahassina@cerist.dz

Ahmed Amine Aliane
Research and Development in Digital Humanities Division
Research Center on Scientific and Technical Information (CERIST)
Algiers, Algeria
aliane781@gmail.com

2022 5th International Symposium on Informatics and its Applications (ISIA) | 978-1-6654-7473-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIA55826.2022.9993553

***Abstract*—Sarcasm is one of the main challenges of sentiment analysis systems. This paper mainly focuses on the recognition of Arabic sarcasm on Twitter. Recognizing sarcasm in tweets is essential for understanding users' opinions on various topics and events. There are only a few attempts regarding saracsm detection in Arabic due to the challenges and complexity of the Arabic language. We propose in this paper a comparison between traditional neural network-based models and pre-trained transformers. The experimental results show that transformers are a promising approach for the task of Arabic sarcasm detection.**

***Index Terms*—Figurative Language, Sarcasm, Natural Language Processing, Twitter, Deep Learning, Arabic Language.**

I. Introduction

User-generated content on social media platforms such as Twitter and Facebook is a valuable source of information for sentiment analysis, which is of great interest to individuals, companies and governments.

Nevertheless, this analysis faces various challenges, especially linguistic ones such as the treatment of negation, dialects, and figurative language. Sarcasm is a particular type of figurative language, which usually shows that the speaker is impatient or contemptuous by saying the opposite of his or her intention in a hostile tone. Sarcasm as a psychological and linguistic phenomenon can be seen as an implicit form of sentiment [1]. Some people use it to react or cope with a reality that may be cruel or a negative situation in society, economy, a political event, or to express an opinion on a controversial topic [2].

Sarcasm detection has received considerable attention in the NLP field. Researchers have begun to focus their efforts on classifying a text (an entire document, a single sentence, a word, etc.) as sarcastic or not. However, this task is difficult and complicated due to the nature of sarcastic texts, which can be relative and differs considerably from person to person depending on factors such as the topic, the region, the event, and the mentality of the user. In other words, a sentence that one person might find sarcastic may seem natural to another.

In addition, the amount of labeled data available for this task is very small, and any available text that can be collected requires a number of human annotators to prepare the data. In addition, detecting sarcasm in the Arabic language poses other challenges, including difficulties in processing the morphologically complex natural Arabic language and the lack of tools and resources available to extract Arabic sentiments from text. This complex task is further exacerbated when dealing with Arabic dialects that do not follow the formal grammatical structure of Modern Standard Arabic (MSA) . [3]

In this paper, we describe models for detecting sarcasm in Arabic. We approached this challenge first by evaluating first traditional neural models (LSTM, Bi-LSTM and CNN) with different word Embeddings and then by experimenting with BERT-based models (MARBERT and AraBERT).