

RUDIMENTARY LEXICON BASED METHOD FOR SARCASM DETECTION

Peter Clews

University of Worcester

UK

p.clews@worc.ac.uk

Dr. Joanne Kuzma

University of Worcester

UK

j.kuzma@worc.ac.uk

ABSTRACT

The purpose of this paper is to establish if rudimentary methods can be used for classifying text as being sarcastic using data taken from the social media website, Twitter. Data collection for this study was carried out using text extracted from Twitter. It applies string matching against positive sentiment and interjection lexicons to test if the presence of both can be used to classify content as being sarcastic. The result shows that the most frequently used terms are too generic to be suitable for a sarcasm specific lexicon. It further shows that Boolean matches to two lexicons are suitable for classification of text as being sarcastic. This is significant as many methods require significant time in collecting and analysing the data to be used within the classification process, as well as complex algorithms to conduct the task. By using simplistic processes, it is hoped that some of the challenges related to performance are overcome. Additionally, this study is the precursor to planned further research into sarcasm detection methods.

Keywords: Sentiment analysis, Twitter, sarcasm detection, positive sentiments, tweets.