PAPER REPORT FOR TASK - 1

Paper Title: Detecting Abusive Bangla Text in Social Media: A Proposed Algorithm and Experimental Results

Paper link: https://ieeexplore.ieee.org/document/8660863

Summary

Motivation:
The paper addresses the challenge of detecting abusive text in Bangla language on social media platforms. With the increasing use of Bangla in social networking, there is a growing concern about cyber oppression, online harassment, and the need for effective detection methods. The study aims to propose a new algorithm for detecting abusive Bangla text and presents experimental results to assess its performance.

Contribution:
The research contributes a novel algorithm for the detection of abusive Bangla text on social media. The proposed algorithm involves a combination of training and testing phases, incorporating features such as unigram, bigram, and trigram. The study emphasizes the challenges in processing Bangla text and the complexity of differentiating between hate speech, funny speech, and abusive speech. Experimental results demonstrate the accuracy of the proposed algorithm for binary classification tasks.

Methodology:
The methodology includes data collection from Facebook pages, Prothom-Alo news, and YouTube channels. A survey is conducted to label comments, and a preprocessing step is implemented to remove special characters, punctuations, and emotions. The proposed algorithm involves training and testing phases, with a focus on feature extraction using unigram, bigram, and trigram structures.

Experimental Results:
The dataset is divided into sets of 100, 200, and 300 comments. Experimental results indicate the correctness of abusive and non-abusive classifications for 20% test data in each set. As the number of comments increases, the accuracy of the proposed algorithm improves, showing promise for detecting abusive Bangla text in social media.

Discussion:
The study discusses future work, suggesting potential improvements such as integrating the algorithm with machine learning techniques, introducing new features, and developing applications for real-time detection. The need for increased dataset size and ongoing efforts to enhance experimental methodologies in natural language processing for Bangla language are highlighted.

Conclusion:

In conclusion, the paper proposes an algorithm for detecting abusive Bangla text on social media and presents encouraging experimental results. The research contributes to addressing the rising challenges of cyberbullying and online harassment in the context of Bangla language. Future work is anticipated to refine the algorithm and explore additional avenues for improvement in the detection of various text types in Bangla.