Paper Title: Cyberbullying detection and classification with improved IG and BiLSTM

Paper Link: https://ieeexplore.ieee.org/document/9996977

Summary:

1. Motivation: Cyberbullying is a growing problem with negative social and emotional consequences. The use of social media is growing, and so are the instances of cyberbullying. Unfortunately, manually identifying and addressing cyberbullying incidents can be challenging. This paper explores the development of an automated system for detecting cyberbullying more effectively.

2. Contribution: This paper introduces a model that utilizes a Bidirectional LSTM architecture with information gain algorithm for feature selection. This approach addresses the challenge of imbalanced datasets in cyberbullying classification and leverages contextual information within sentences to improve accuracy.

3. Methodology: They used a more improved information gain algorithm since standard information gain can be biased towards datasets with imbalanced categories. Futhermore, the model employs various preprocessing techniques like lemmatization to bring back a word to its natural form. It also removes stopwords and emojis which likely improves the model's ability to focus on the semantic meaning of the text.

4. Conclusion: The proposed model achieves high accuracy in identifying cyberbullying and its category on Twitter comments reaching overall precision of the model at 95.15%.

Limitations:

Although the model provided in this research shows promising results, there are still issues that need to be addressed further. The model solely relies on the content on twitter comments for classification. It's unclear how well the model will perform in other social media platforms with different language styles or types of cyberbullying that might not be common on twitter. While the paper explores the effectiveness of the proposed model with various configurations, the comparison with other models is limited. The evaluation primarily focuses on Naive Bayes models and a Unidirectional LSTM with the information gain. A more detailed understanding of the Bidirectional LSTM's performance in comparison to other models, such as logistic regression or KNN, would be beneficial.

Synthesis:

This paper presents a well-structured model with promising results for cyberbullying detection on Twitter comments. The combination of data preprocessing, information gain for feature selection, and a Bidirectional LSTM architecture demonstrates effectiveness. As the paper suggests, incorporating user information beyond comment content is a promising avenue for future research. Additionally, exploring other deep learning architectures like transformers or convolutional neural networks (CNNs) could potentially improve performance. Evaluating the model's generalizability to different platforms and cyberbullying types would also be valuable. Moreover, exploring alternative evaluation metrics that consider the severity and cost of misclassification would provide a more comprehensive understanding of the model's effectiveness in real-world applications. Overall, this paper presents a valuable contribution to the field of cyberbullying detection, laying the groundwork for further advancements in automated systems to combat this critical issue.

ID: 21201357
Zakaria Ibne Rafiq