Paper Title: Detection and Fine-Grained Classification of Cyberbullying Events

Paper Link: https://aclanthology.org/R15-1086.pdf

Summary:

1. **Motivation:** The increase use of social media has made cyberbullying an issue that has to be addresses by the users. The report emphasizes on the automatic identification of cyberbullying and investigate the viability of identifying the more specific types of cyberbullying in internet postings.

2. **Contribution:** The paper uses ML algorithms to detect textual forms of cyberbullying which was a binary classification (cyberbullying versus non-cyberbullying). The research used two different ways to approach the detection, first a rule-based model and secondly a bag-of-words model. After the experiment if seemed that the role-based model out-performed the bag-of-words model as the role-based model used lexical features. The experiment also got a helping hand from the experts and the users to get a better grasp of the detection process.

3. **Methodology:** The data was fetched form a social networking site Ask.fm where the identity was not disclosed. The data was taken in question and answer based form. The annotation was based on two different level. Firstly, the annotators were instructed to mark if a post is a part of an incident of cyberbullying at the post level. This was carried out giving the post a harmfulness on a three-point scale: 0 represents no signs of cyberbullying, 1 indicates mild indication of cyberbullying and 2 indicates serious indications of cyberbullying (e.g., physical threats or incitements to commit suicide). Secondly, the annotators identified finegrained text categories related to cyberbullying at the subsentence level, identifying all text spans corresponding to one of the categories described in the annotation scheme.

4. **Conclusion:** The paper presents promising results for identifying cyberbullying events and fine-grained classifying insults using lexical features. The study aims to reduce manual monitoring effort on social media and explores recall optimization and author role information for enhanced detection. The researchers plan to use advance features like synthetic patterns and semantic information to recognize implicit realization of cyberbullying. They also examine features selection technique to reduce vector sparseness and orthographic normalization of data. Although the experiment was performed on a Dutch dataset it can be constructed on a English dataset

Limitations:

Limitations that can be mentioned are:

a) The primary topic of this paper is the identification of cyberbullying on textual form and it might not function on other systems.
b) Collections of data was performed on a singular platform and the construction for other languages differ alongside with the classifications.

Synthesis: The paper gives more priority to the posts regarding insults and curses alongside with the threats in a textual form but how ever there are more than one form of cyberbullying occurring in this world and the advancement of technology grows. To overcome the other forms the ML algorithms that were in use to detect the threats are not suitable. The real-based approach or a bag-of-words approach cannot be implemented on image form of cyberbullying. Here the algorithms that were in used to test the detection were good but not as up to the mark of the recent days used algorithms which can detect textual cyberbullying with more accuracy and precision. The paper informs us that using lexical features are more efficient it gives the researcher to have a strong base line to improve on the model.