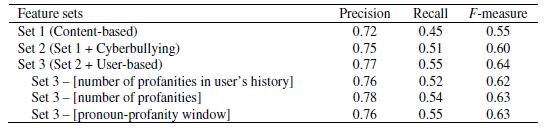
Improving Cyberbullying Detection with User Context

Abstract. The negative consequences of cyberbullying are becoming more alarming every day and technical solutions that allow for taking appropriate action by means of automated detection are still very limited. Up until now, studies on cyberbullying detection have focused on individual comments only, disregarding context such as users’ characteristics and profile information. In this paper we show that taking user context into account improves the detection of cyberbullying.

Dataset:   
4626 comments from 3858 distinct users. The comments were manually labelled as bullying (9.7%) and non-bullying (inter-annotator agreement 93%).

Methodology:

1. Feature Space Design
   1. Content-Based Features {set1}
      1. Normalized number of profane words in the comment based on a dictionary
         1. http://www.noswearing.com/dictionary
      2. Normalized number of first and second person pronouns
      3. Profanity windows of different sizes (2 to 5 words)
      4. The number of emoticons was counted and normalized
      5. Ratio of capital letters in a comment was computed
   2. Cyberbullying Features {set2}
      1. Number of cyberbullying words
      2. The length of the comment.
   3. User-Based Features {set3}
      1. History of user’s activities to see whether there was a pattern of offensive language use and frequency of profanity in their previous comments.
   4. As a preprocessing step, stop-word removal and stemming were applied
2. Experimental Setup
   1. Support Vector Machine

Drawbacks:   
The low recall of the first feature set can be explained by the occurrence of bullying comments without explicit profanities and by implicit bullying through sarcasm. The age feature had contributed but not as much as expected in the classification of bullying comments. The latter might be due to the fact that many users do not indicate their real personal information.