Bullying Analysis Using Twitter Data

Hoanh Nguyen
University Massachusetts Lowell
1 University Ave
Lowell, MA 01854, USA
soujiroboi@gmail.com

Michael Meding
University Massachusetts Lowell
1 University Ave
Lowell, MA 01854, USA
mikeymeding@gmail.com

Abstract

Twitter is a social network where users can communicate publicly with short text statements called tweets. Since it began, the way Twitter is used has changed dramatically. It has gone from a calm social environment to a much more surprisingly hostile place to connect with others. For this reason the program that searches twitter for people who bully those around them. We did this by taking data from twitter and analysing the aggressiveness of given tweets using the Senti-WordNet database and improved using Harvard Inquirer. This is then put into a Machine Learning algorithm to try and predict whether a users posted tweet can be classified as bullying.

1 Introduction

This project is a program that is used to identify twitter users who are bullies to those around them. This might seem straightforward at first but looking at this problem in more detail it is actually quite complex. This program trains a binary classifier to decide if a single tweet is bullying or not. Unfortunately, bullying is not a binary issue and therefore is quite difficult to discern given that tweets are so short. Often times the subject of an aggressive tweet is implied making it very difficult for a machine to find a subject. The solution to this was an assisted machine learning algorithm using a brute force approach. Given good annotated data to begin with the results were surprisingly good given small data sets.

2 Related Work

The University of Wisconsin did a project two years ago on the study of bullying in twitter that heralded lots of media coverage including several articles from Huffington Post and Time Magazine. The first version of their code was made public that same year and suffice to say that this project has improved on their initial algorithm. Their code used a set of static words as search terms for tweets then classified by a very small existing bullying model. This model did yield some results but the efficacy left much to be desired.

3 Methodology

3.1 Machine Learning Methods

3.2 Data Sets

The data for this project was gathered by a data mining algorithm. The algorithm uses a lexicon of twitter hashtags and words that have been tagged with a sentiment score. Based on the words which have medium to high negative sentiment score it then retrieves the most recent tweets with that word. It then filters the received tweets for spam and for relevancy before compiling into a corpus file. Based on the aforementioned level of sentiment score for given search terms it is able to control the size of the output file as individual search terms are limited to 40 tweets per query.

3.3 Evaluation

4 Results

5 Future Improvements

None. We kicked Twitters ass.

Acknowledgments