

Comment Abuse Analyzer

CLINTON PAHL, JAMES OSGOOD, PAUL VANDERMEER*

University of Alberta

cpahl@ualberta.ca osgood@ualberta.ca pv1@ualberta.ca

Abstract

Since the world is depending more and more to computers to feed people information there tends to be a chance that there will be more and more abuse posted. Not everyone wants to see these messages as some of them could be profound and very abusive to some. Our project is going to predict whether there is abuse detected in a publicly posted message. To solve this problem we have tried a couple of different methods.

I. PROBLEM

Our project entails detecting abuse in public online comments. This is a common problem that is getting worse each and every day. For website administrators or moderators to manually go and moderate all the comments on their website is very costly and time consuming. Since performing this task is time consuming a lot of stuff does get missed or does not ever get looked at. If every comment was to be looked at first posts would take a long time to actually hit website and people would be turned off and go elsewhere. The big problem we are faced with is whether certain comments are considered abuse or if they are just regular comments. Since profanity can be used in a regular unobscene way the training data used needs to be refined so that different comments are classified differently.

II. METHODS

The methods we are going to investigate are: Bootstrapping, Bag of Words, String Kernel. We chose these methods as we hypothesized that it would be easier to parse out the features we require. Here is just a summary of what these methods entail.

Bootstrapping

Lets make a dataset D_n where n is the number of samples in the dataset. Then we take $j = 1 \rightarrow k$, where k is how many samples we are going to use for bootstrapping. This new dataset will be chosen by random where the j_{th} entry will replace the current entry in D_n . After this process is done we will end up with a modified dataset D_n that has some entries in duplicate and it may also have removed some items.

Bag of words

With the bag of words algorithm we take all the words of the training set and uniquely place them in a new ordered set. The words in the phrases now create a matrix where each row is created by a method that introduces a weight count for each time the word exists. This special method called "Term weighting" uses hashing to efficiently compare which words in the comment are in our dictionary and thus limits the resources required to run our algorithm.

*A thank you or further information

III. LITERATURE SURVEY

pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330.

IV. APPROACH

V. RATIONAL

VI. HYPOTHESES

VII. EXPERIMENTAL DESIGN

VIII. RESULTS

IX. CRITICAL EVALUATION

X. LESSONS LEARNED

REFERENCES

[Figueredo and Wolf, 2009] Figueredo, A. J.
and Wolf, P. S. A. (2009). Assortative