Troll Project

Matt Cole

October 7, 2016

Introduction

Internet trolls are considered a menace in nearly all online communities by creating fruitless arguments in an attempt to generate emotional reactions from comments. Motivations behind these users are unclear, but some social scientists have postulated that a strange phenomenon known as the "disinhibition effect" may be to blame. Masked behind the apparent anonymity provided by many forms, sites, and the internet in general, social reservations that facilitate normal, face-to-face conversations can disappear, resulting in sometimes wild and rude behavior. Troll behavior can stretch from simply posting the same status repeatedly to annoy and clog streams of information to violent threats and many, many 'things' in between.

In this project, we focused on political trolls on Twitter, those whose tweets aim to disrupt the flow of information in twitter's political sphere.

Data

Trolls were defined as users with tweets which appeared to both:

- inflammatory: intended to arouse angry or violent feelings
- extraneous: irrelevant or unrelated to the subject being dealt with

Data was collected from twitter using both the Twitter API and web app. Initial data collection consisted of both searching twitter for political figures in tweet text (ie. clinton, trump). Manual user harvesting consisted of browsing political figures' accounts in an attempt to find associated users whom were and manual troll harvest. In total, around 2000 tweets from 271 individual accounts were assessed and categorized by reading the most recent 10 tweets at time of data collection from each user.

Using data from the last ten tweets (inclusive of retweets), we were able to assess average tweet sentiment using the AFINN lexicon which comprises of English words with valence measures between -5 and 5 inclusive with an integer between minus five (negative) and plus five (positive). Additional lexicons which include the NRC Emotion Lexicon, which assigns 'scores' to each of two sentiments and eight emotions were used too. As emojis have become increasingly popular in today's web-based communication, an emoji dictionary was scrapped from unicode.org and emoji usage as well as sentiments was assessed.
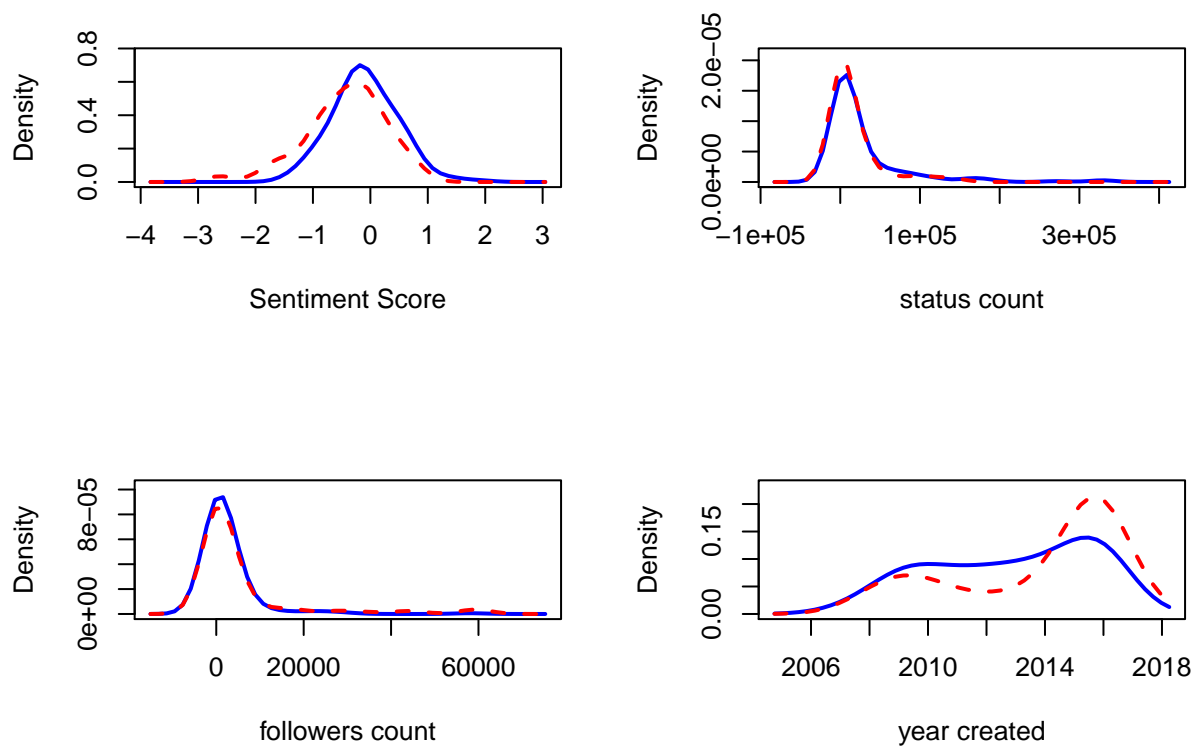
Several statistical and machine learning techniques were used in an attempt to predict whether a user was a troll or not including linear regression, cart, and svm.

Results

Pruned Classification Tree for Troll
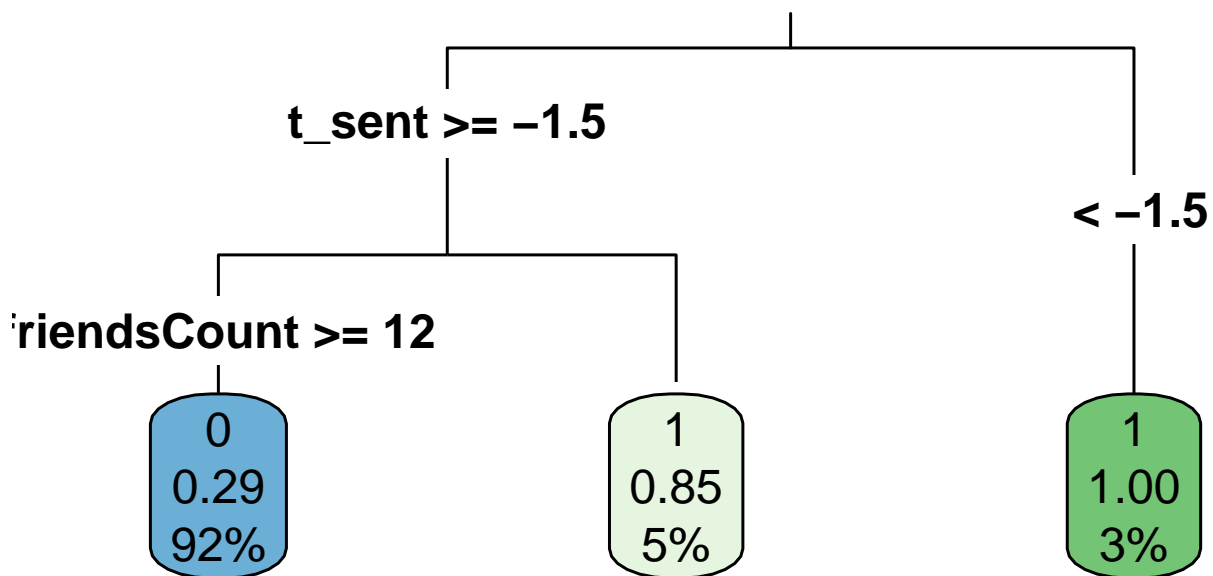

```
## [1] 0.2105571
```

```
## [1] 0.2035869
```

Here we see that the pruned regression tree has a cv-prediction error rate of `r` using a simple logistic regression we found a prediction error of 0.31 and including more relevant variables 0.28. The regression tree was pruned according to 10-fold cross vadiated prediction error.

Discussion

- regression trees ** recursive partitioning

Thank you

Finn Årup Nielsen manually labeled the words in the AFINN from 2009-2011. Thank you to http://unicode.org/emoji/charts/full-emoji-list.html for letting me borrow the list.