Social Media Moderation with Sentiment Analysis

Introduction

Reddit is one of the internet’s most popular online communities where users post and share information, specifically in the form of comments. Users vote posts and comments up or down based on their opinion and earn “karma” for their participation. Reddit is made up of various communities known as sub reddits, centered around specific topics.

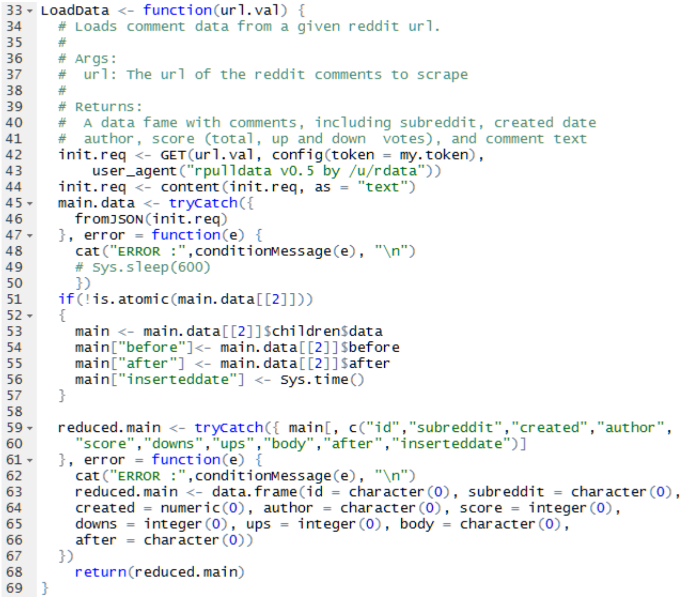
Popular posts or comments, earn “karma” but not all popular posts contribute to the community. If there are too many negative posts in a community people may be disincentivized to share and Reddit will lose users. Although each post has a “score” based on up and down votes, there is not a classification system that categorizes comments based on their overall sentiment.

This project performs sentiment analysis on a sample of comments from Reddit and develops an algorithm to identify correctly classify comments “positive”, “neutral” or “negative”. It is useful to identify users and communities who have consistently have negative posts. By classifying comments by sentiment and identifying negative posts, reddit can ensure that it removes or flags users or comments who are consistently destructive to the community and rewards users who provide positive participation, not just those who are the most popular.

Data Set

Data was collected from the reddit API outlined at reddit.com/dev/api using endpoint https://oauth.reddit.com/r/all/comments/.json. Using an oAuth2 connection as a registered user application, we were able to collect over one million comments from almost 17,000 subreddits spanning a sample four-day period, 5 July to 8 July 8 2016.

Using rCurl, jsonlite, httr and httpuv libraries the relevant information was downloaded in json format and then converted the data to data frames. With r.utils and DBI data was saved to a SQL database.



```
LoadData <- function(url.val) {
  # Loads comment data from a given reddit url.
  #
  # Args:
  #   url: The url of the reddit comments to scrape
  #
  # Returns:
  #   A data fame with comments, including subreddit, created date
  #   author, score (total, up and down  votes), and comment text
  init.req <- GET(url.val, config(token = my.token),
      user_agent("rpulldata v0.5 by /u/rdata"))
  init.req <- content(init.req, as = "text")
  main.data <- tryCatch({
    fromJSON(init.req)
  }, error = function(e) {
    cat("ERROR :",conditionMessage(e), "\n")
    # Sys.sleep(600)
    })
  if(!is.atomic(main.data[[2]]))
  {
    main <- main.data[[2]]$children$data
    main["before"]<- main.data[[2]]$before
    main["after"] <- main.data[[2]]$after
    main["inserteddate"] <- Sys.time()
  }

  reduced.main <- tryCatch({ main[, c("id","subreddit","created","author",
    "score","downs","ups","body","after","inserteddate")]
  }, error = function(e) {
    cat("ERROR :",conditionMessage(e), "\n")
    reduced.main <- data.frame(id = character(0), subreddit = character(0),
    created = numeric(0), author = character(0), score = integer(0),
    downs = integer(0), ups = integer(0), body = character(0),
    after = character(0))
  })
    return(reduced.main)
}
```

Figure 1 Code to Load Data Set from Reddit API

The important fields in the data set are:

Field Name	Description	Example
id	Unique identifier per comment	d4z5y3s
subreddit	Community comment originated	AdviceAnimals
created	Time created (UTC Timestamp)	1467691477
author	User name of commenter	eric881
downs	Number of negative down votes	0

ups	Number of positive up votes	1
body	The text of the comment	I'm new to basketball trades and stuff, but is it just a small group of incidents that players go to other teams? Or are there multiple times players trade?
after		t1_d4z5y3s

In addition to the existing fields, several fields were added in order to help with the model and as part of the data cleaning process:

Field Name	Description	Example
score	Sentiment score using a basic integer score algorithm that subtracts negative words from positive words using an opinion lexicon	0
inserteddate	Time comment was inserted into the database	2016-07-04 16:04:37.693
score_category	“Positive”, “Neutral” or “Negative”	Neutral
created_time	Time created (date time format)	2016-07-05 04:04:37.000

The data was cleaned to remove unnecessary characters using gsub() and a basic regex and conversion to lower case, as well as using strsplit from stringr.

The data was split into a training set (60%), a validation set (20%) and finally tested on new data (20%). Each comment in the training set was assigned a sentiment score based on the sum of the words used using a Hu Liu’s opinion lexicon.

Limitations

Some of the interesting questions that cannot be answered include location based information or other demographic based information. Because of the anonymous nature of the Reddit community there is no way to collect this type of information linked to specific individuals. Information by region or country may have been interesting.

Additionally, a large number of comments may have been posted by “bots” or automated programs designed to automatically post comments in a predictable pattern when triggered by a post, often based on key words. This may skew the results.

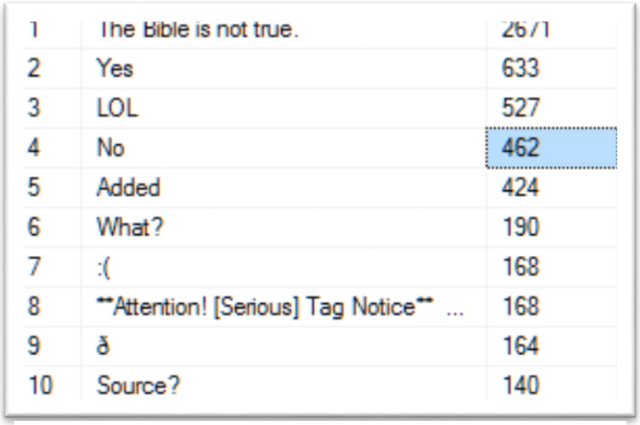
Initial Findings

Most comments (77%) are clustered around a score of -1, 0, or 1, with a plurality with a score of 0 (43%). This suggests that most comments do not present strong sentiments on either the positive or negative side.

Additionally, preliminary data exploration shows that not all comments are posted in English, as well as the common usage of “shorthand” such as using “k” for “ok” which may make the classification process less accurate.

Approach

A prediction model will be developed to help solve the classification problem focusing on naïve Bayes classification. The training set will be classified in with a simple, well established approach using an existing lexicon.



1	The Bible is not true.	2671
2	Yes	633
3	LOL	527
4	No	462
5	Added	424
6	What?	190
7	:(	168
8	**Attention! [Serious] Tag Notice** ...	168
9	ð	164
10	Source?	140

Figure 2 Sample of frequent short comments

The initially proposed approach still applies; however, a much smaller data set will be used. Initially we wanted to use comments over a year time period. This presented two problems – obtaining a dataset of past comments is difficult using the Reddit API and the sheer volume of data would have been unwieldy to process. The gathered data set of over one million unique comments was gathered over a one-week period.