Sentiment analysis over “gun control”

1). Introduction

“Gun control” is always is topic in US. Recent mass shootings in America reveal the serious of this problem. I wanted to see people’s opinion above “gun control” through sentiment analysis.

2). Literature Review

There are some articles online mention about different approach used for “twitter sentiment analysis”. While some use automated sentiment analysis solutions such as “monkeylearn”, other use Python or R. Basic they all label dataset with positive / negative / neutral rating and use them to train a model.

3). Dataset

I’m going to download about 1000 tweets through Twitter’s REST API, because I can’t find any public dataset focus on “gun control”. I’m going to label 200 of them manually, then create a model to classify them all. Only attribute I need is text of each tweet.

4). Approach

Step 1: Download twitter datasets

Login the Twitter Developer website and create an application, follow step on <http://rtweet.info/articles/auth.html>. Then scrape tweets against hash tags “guncontrol” and store the tweets into a csv. I already downloaded the dataset.

Step 2: label the tweets and apply NLP techniques to extract keyword

Label 300 tweets as with positive or negative, use TF-IDF method for text preprocessing and “Word cloud” to extract the keywords for “For” class and” Against” class.

Step 3: Train tweets using different classification approaches

Using 300 classed tweets as dataset then apply KNN, Decision Tree and Naive Bayes to create the classification model. Compare three approaches for highest accuracy.

Step 4: Apply the best model to twitter datasets

Use the highest accurate model to label the rest of 1700 tweets and get result of analysis.

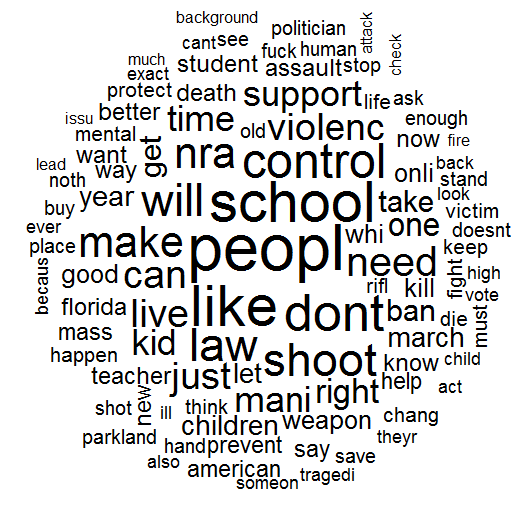
Link to GitHub repository:

https://github.com/seanhuang88/Sentiment-analysis-over-gun-control-

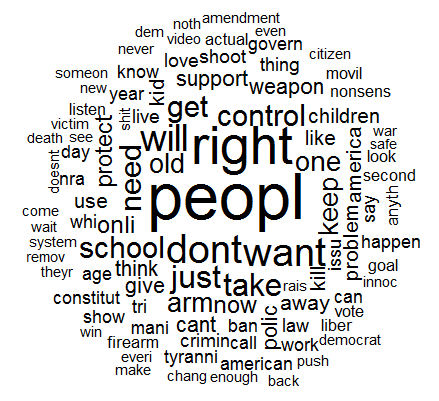
5). Result

A. World Cloud

I found as show below that people support “Gun Control” talk about “People”, “School” mostly.



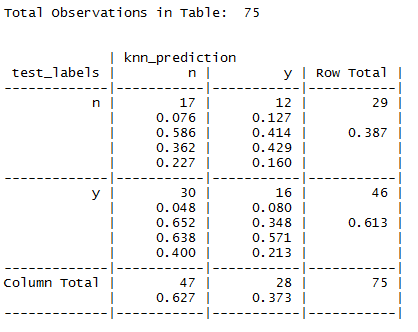
And people against “Gun Control” talk about “People”, “right” mostly.



There are some common words used by both side as “People” “Don’t”, also I already excluded “Gun”.

B. KNN, Decision Tree and Naive Bayes comparison

KNN: (confusion matrix)



Accuracy: 0.1484099

Decision Tree: (confusion matrix)



Accuracy: 0.2261484

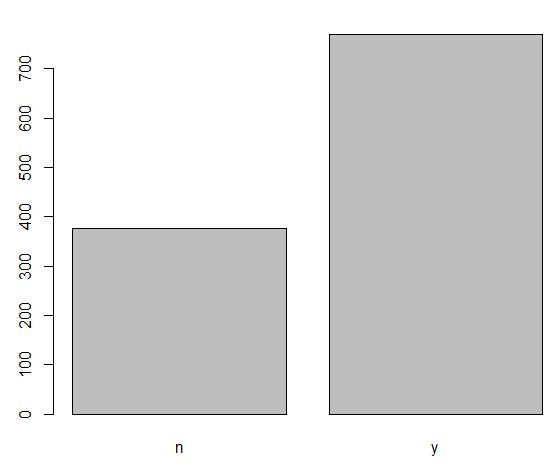
Naive Bayes: (confusion matrix)



Accuracy: 65.33333

It seems Naive Bayes has highest accuracy.

C. Final result



Apparently there are much more people support "Gun Control" in US.

6). Conclusions

Result from twitter sentiment analysis shows that currently more people support "Gun Control", but I also found with time passing by, more and more people stand up to fight their right of owing gun. I pulled 1000 tweets in two different period, the latest data show trend of favor right of self-defense.

Lots of people did twitter sentiment analysis, but few mention about the difficulty of collecting and cleaning the tweet, the dataset I got by searching hashtag “Guncontrol” include many irrelevant tweets and also you need to have extensive knowledge about what’s really happened in US to be able to correct label the dataset. Also as I did, the more tweets you have the more accurate result you will get. It's quite time consuming to manually label the training and testing tweets. For real organizational project, I believe we need at least tens of thousands of tweets and repeatedly follow-up in order to gain an accurate result.

REFERENCES:

(1). Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach

https://www.ijariit.com/manuscripts/v2i5/V2I5-1157.pdf

(2). Analyzing Twitter Sentiment of the 2016 Presidential Candidates

https://web.stanford.edu/~jesszhao/files/twitterSentiment.pdf

(3). Sentiment Analysis on Twitter Data using KNN and SVM

https://thesai.org/Downloads/Volume8No6/Paper\_3-Sentiment\_Analysis\_on\_Twitter\_Data\_using\_KNN\_and\_SVM.pdf

(4). Twitter sentiment analysis with Machine Learning in R using doc2vec approach

https://analyzecore.com/2017/02/08/twitter-sentiment-analysis-doc2vec/