1. Describing the Problem

1.1 Problem

The problem is ranking a Tweet's sentiment. This is a very basic use case. The goal is to define and implement this method using research and libraries on convolutional neural networks. This project will demonstrate how by labeling a Tweet's sentiment, one can begin to categorize a Twitter account.

1.2 Problem Type

The type of problem is sentiment analysis, and classification. We will perform sentiment analysis by processing the text and classifying tweets on a scale from -1 to 1; -1 being very negative and 1 very positive.

2. Background

2.1 Project Motivation

The motivation for this project to determine if a Tweet is positive or negative and utilize that label to rank account's on a linear, one-dimensional, sentiment scale. People reading the tweet can be unsure of the Tweet's connotation, but we can statistically estimate how positive or negative it is. Twitter is one of the largest publicly available datasets, and therefore provides a perfect avenue for creating a model to detect sentiment on a microblogging platform. Use cases extend to banning users who produce hateful speech, detecting users associated with terror/hate groups, or simply removing hateful content to produce a friendly platform. (10)

2.2 Background Resources

2.2.1. Literature Resources

This project can be approached as one that has a solutions in machine learning, natural language processing, artificial intelligence and sociology. Therefore the sources and publications considered as background for this project will contain topics pertaining to one or more of these fields.

2.2.1.1 Machine Learning

Research publications pertaining to this topic were helpful in providing avenues for training data, methodologies for testing algorithm accuracy, and motivation for performing this research. These publications also provided interesting use cases, for instance detecting extremist supporters (2).

2.2.1.2 Neural Networks

The bulk of our background research pertains to this topic. From these publications we gained insight into the structure of the data, semantic methods to classify tweets, readily available training data, how to construct a convolutional neural network, and how to test its accuracy. Important lessons to note are that: “Sentiment analysis of short texts such as single sentences and microblogging posts like Twitter messages is challenging because of the limited amount of contextual data in this type of text” (6) and therefore we must “...go beyond bag-of-words and extract information from the sentence/message in a more disciplined way” (6). Therefore “...to fill the gap of contextual information in a scalable manner, it is more suitable to use methods that can exploit prior knowledge from large sets of unlabeled texts” (6). This specific text “...propose[d] a deep convolutional neural network that exploits from character- to sentence-level information to perform sentiment analysis of short texts” (6)

3. Describing the Data

3.1 Data Kinds

The data will be in a very generic format. Once process and parsed, they will be Tweets with removed metadata features like emoticons, usernames, hashtags, links etc.

3.2 Data Obtainment

3.2.1 Pre-Trained Word Embeddings (Unsupervised)

There are two data sources we will motivate for pre-trained work embeddings. These are the GloVe word vectors, which are pre-trained on 2 billion Tweets published by Stanford NLP group.

The second is the Word2Vec, an algorithm which produces shallow 2 dimensional neural networks using skip grams, because the emphasis will be on words which are closer in context due to the short length of tweets.

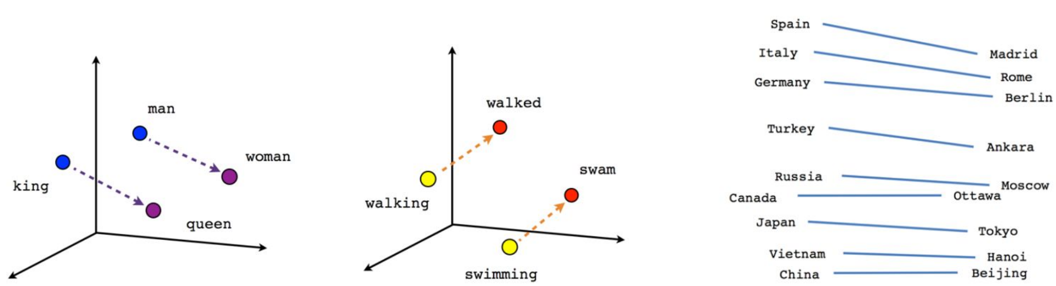
3.2.2 Labeled Training Data

The SemEval project has committed their research efforts to annually producing labeled training data for Twitter sentiment analysis. We will leverage this already existent, open source, pre labeled data to train our convolutional neural network. A statistic properly to note of the labeled training data is that it is skewed toward a more positive sentiment. Meaning there is more data labeled as positive than there is negative. Smoothing solutions will be discussed later.

3.3 Data Features

3.3.1 Unsupervised Features

The unsupervised dataset will produce a word vectors and plot them in a multidimensional matrix based on multiple properties. The following figure is a diagram of plotted word vectors.



Male-Female Verb tense Country-Capital

3.3.2 Labeled Features

The labeled data consists of a processed Tweet, removed of its metadata through the methods described in this proposal, and an indication of whether the Tweet is positive or negative.

id1	id2	start_token	end_token	sentiment	tweet_text
19293810	15115101	3	4	positive	who's a master

3.3.3 Testing Features

We will test our model on the labeled training data from a separate year and this will contain the same labels as above. We will also test our data on manually observed Twitter API random stream data and see how the model categorizes them.

4. Outcome

4.1 Outcome Computed

Outcome will be a table of columns indicating a text of tweet and a score from -1 to 1 indicating its affinity. Sample Figure:

Tweet	Score
“I passed the test”	1
“I failed the test”	-1
“The test was easy, but I failed”	0.25
“The test was hard, but I passed”	-0.25

4.2 Outcome Type

The type will be a categorization of a Twitter account by the percentages of its sentiment. These types of observations have not been performed before. Although current research has implemented convolutional neural networks to perform sentiment analysis, none have observed the sentiment percentage of Twitter accounts.

4.4 Outcome Usefulness

This kind of outcome can be used to remove hateful speech from Twitter by removing accounts that statistically produce hateful content on a regular basis.

4.5 Confounders

There are several confounders for this project. Some are attitude, day of the week, or environmental circumstance. Just because the Tweet's content is a specific sentiment, does not mean that an overall account or persona contains that specific sentiment without taking into consideration these other confounders.

5. Describing the Methods

5.1 Model

5.1.1 Convolutional Neural Network Model

Neural networks are prone to overfitting, so we will put an emphasis on regularization to generalize the model for unseen data sets. We will implement a convolutional neural network. Convolutional Neural Networks "overfit on small and medium sized datasets" (9) so current research suggests that we "oversize the cost function with l2-norm regularization terms for the parameters of the network" (9).

5.1.2 Accuracy/Testing

We will measure accuracy using an array of testing methods. These are with a confusion matrix (TP, TN, FP, FN), precision & recall metrics, ROC curves, and a 10 fold-cross-validation.

Also can test the accuracy of our models compared to baseline models like naive bayes (4).

5.2 Approach

We will train the convolutional neural network on both a feature matrix (the word vectors) and the pre-labeled training data. This will provide both a contextual understanding of Tweets and a natural language processing understanding as well.

5.3 Assumptions

The main assumption here is with the pre labeled training data set. We assume that the participants who labeled this data were conscious enough of overall sentiment to accurately label the Tweets as positive, negative, or neutral.

6.1 References

1. https://arxiv.org/pdf/1605.00659.pdf
2. https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf
3. https://arxiv.org/ftp/arxiv/papers/1509/1509.04219.pdf
4. http://cs229.stanford.edu/proj2013/Tai-TwitterSentimentAnalysis.pdf
5. https://cs224d.stanford.edu/reports/YuanYe.pdf
6. http://anthology.aclweb.org/C/C14/C14-1008.pdf
7. https://zhangmeishan.github.io/Ren.pdf
8. https://www.aaai.org/ocs/index.php/AAAI/AAAI13/paper/download/6419/6821
9. http://delivery.acm.org/10.1145/2770000/2767830/p959-severyn.pdf?ip=108.41.226.114&id=2767830&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=683567419&CFTOKEN=16058570&__acm__=1476892666_0701480dd0a146463fe42b21d2caab70
10. http://qz.com/812314/the-hate-speech-on-twitter-twtr-reportedly-steered-disney-dis-away-from-buying-the-company/

6.2 Repositories

- https://blog.rstudio.org/2016/09/15/tidyverse-1-0-0/
- https://github.com/Lab41/sunny-side-up/wiki/Deep-Learning-Techniques-for-Sentiment-Analysis
- http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/
- https://github.com/xiaohan2012/twitter-sent-dnn
- http://www.fredericgodin.com/software/