Zack Godwin

A Sentiment Analysis RNN

16th January 2019

OVERVIEW

Brand management is a key factor for a successful business. Understanding how customers feel about a brand, company, or even a specific topic can help predict how they'll respond to future decisions. Twitter gives us the perfect data stream to investigate this.

GOALS

1. Stream at least 200,000 tweets from Twitter using the Twitter API into a Postgres server.
2. Use the TextBlob sentiment analysis module to label the tweets.
3. Investigate the dataset using NLP, cluster analysis, and EDA.
4. Build a classification model to predict sentiment.
5. Train a RNN classifier and compare performance..

SPECIFICATIONS

I will use the Twitter API along with the Tweepy python module to collect 200,000 tweets in a Postgres SQL database. The Tweets will be filtered to include the word "trump", the idea being that in today's political climate there is both an abundance of tweets regarding Donald Trump as well as stark polarity in how people feel about him. We need the volume to allow for quick data collection, and the polarity for labeling tweet sentiment. Using Word2Vec and Doc2Vec NLP techniques I'll train some supervised machine learning models to classify tweets as either positive or negative. The best one will be compared to an LSTM RNN trained on the same corpus, but without Word2Vec. All coding will be done in the Google Colaboratory environment, and I will leverage Pytorch to develop the Neural Network.