Sentiment Analysis on First Person Narratives

Contributors: Bryan Tor, Harsh Karia, and Sahithi Narla
{btor, hkaria, snarla}@ucsc.edu

INTRODUCTION

Individuals writing informally from the first-person perspective typically express emotions indirectly by describing situations from which their audience will infer a general sentiment, instead of explicitly stating the emotion that the individual is feeling. In 2017, the Natural Language and Dialogue Systems Lab (NLDS Lab) of UC Santa Cruz created a model to predict the sentiments of first-person narratives by learning lexio-functional patterns.

Our group wanted to see if a LSTM model could replicate similar results to the NLDS model, substituting the AutoSlog-TS pattern learner used for the hidden state cell used by a LSTM.

As our group progressed with our project, we experimented with a variety of different models to attempt to replicate the results achieved by the NLDS model.

DATASET AND PREPROCESSING

Our data was obtained from Natural Language and Dialogue Systems Lab (NLDS) at UC Santa Cruz. The data was split into a training set made up of 46,255 positive and 25,069 negative sentences for a total of 71,324 sentences, a testing set made up of 1,266 positive and 1,440 negative sentences for a total of 2,706 sentences, and a dev set composed of 498 positive and 754 negative sentences.

In the NLDS model, the dev set was used to tune the parameters of an AutoSlog-TS pattern learner. Because we were not taking a similar approach, we combined the training set and dev set.

Furthermore, we were unconfident in how the NLDS split their data between a training set and testing set, given that the ratio between the training set and testing set was about 3240 : 120. Our group concluded to build and evaluate our models in two ways: train off of the train set and test off of the test set; merge the test set and train set, and apportion the merged set into new train and tests set by a 66-33 split.

We also noted an imbalance between the number of positive and negative sentences, which will be discussed later.

id	sentence	sentiment
50f70518d7eed88dd8f1de63_2.txt	I remember having an odd dream, but not the whole thing.	0
5101cabcd7eeef2893914552_31.txt	ride my bike maybe?	1
5344221be4b0e2a60cea917d_215.txt	In contrast to what has been written here I am a very optimistic person.	0
50f7476fd7eed88dd8f3072c_11.txt	I can only say and understand a few things, especially the Sono gemeli?	1
50ff4686d7ee253f31855b6c_7.txt	We're basically going to use photoshop and dreamweaver to build a website.	1
53441681e4b0e2a60ce40581_185.txt	I needed you more than anyone else needed you because... news flash...	0
50ff8a69d7ee253f3188413f_11.txt	I don't get all the Oasis haters out there.	1
51030986d7eeef2893967966_8.txt	i think alonso, and massa are good, though massa lost cause of the stupid automated lollipop man and that his team of r	1
50f65fafd7eed88dd8eceee2_1.txt	I suppose people go about securing their critical needs differently when a hurricane threatens.	1
51016475d7eeef28938c62b0_16.txt	u tell me u find out last minute about the sia stuff so u couldnt tell me i go out with my friends and u talk	0

Figure 1. Sample of our Dataset

We wanted to eliminate any features that might be irrelevant for our training our model. We lowercased our data, removed stopwords, removed punctuation and extra whitespace, and lemmatized words.

	Unnamed: 0	id	sentence	sentiment	sentenceProcessed
0	0	50f1ba41d7ee5a6cc987b20e_10.txt	She filled me in and then we were silent as we...	0	filled silent watched news coverage
1	1	50f1ba41d7ee5a6cc987b20e_11.txt	After pulling my wits together, I hung up on h...	0	pulling wit together hung started calling round
2	2	50f1ba41d7ee5a6cc987b20e_16.txt	[By then my folks were living in southwestern ...	0	folk living southwestern virginia four hour dr...
3	3	50f1ba41d7ee5a6cc987b20e_17.txt	We shared some flabergasted moments and then t...	0	shared flabergasted moment report came pentagon
4	4	50f1ba41d7ee5a6cc987b20e_18.txt	I started shrieking.	0	started shrieking
5	5	50f1ba41d7ee5a6cc987b20e_2.txt	As usual, I chatted with the moms and other na...	0	usual chatted mom nanny walked back house
6	6	50f1ba41d7ee5a6cc987b20e_20.txt	He had to hear me fall into a panicked mess.	0	hear fall panicked mess
7	7	50f1ba41d7ee5a6cc987b20e_21.txt	The home where I lived was close to the Dulles...	0	home lived close dulles airport
8	8	50f1ba41d7ee5a6cc987b20e_22.txt	Incoming and outgoing flights went over our ne...	0	incoming outgoing flight went neighborhood reg...
9	9	50f1ba41d7ee5a6cc987b20e_26.txt	She wanted me to hop in the car that instant a...	0	wanted hop car instant drive home
10	10	50f1ba41d7ee5a6cc987b20e_27.txt	Then onto my employer.	0	onto employer

Figure 2. Sample of our Dataset after Cleaning

MODELS

We experimented with five different models:

- Basic SentiWordNet: Assigns a positive or negative value based off of the average sentiment of each word of the sentence according to SentiWordNet. This was used as our baseline model.
- Multinominal Naïve Bayes Classification: Used the Bayes classification algorithm from scikit-learn. Used a Bag of Words and TF-IDF vectorizer when testing.
- Logistic Regression: Used the Logistic regression algorithm from scikit-learn with a TF-IDF vectorizer
- Word2Vec: We obtained word embeddings of our vocabulary using the Word2vec model, and then processed that through a densely-connected neural network layer.

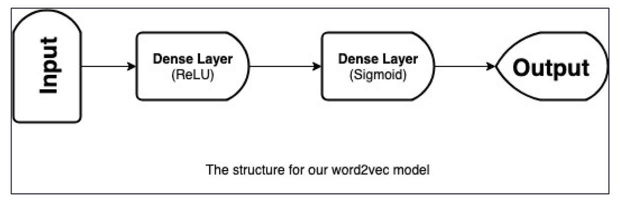

Figure 3. Word2Vec Model

- Recurrent Neural Network: Used a bidirectional LSTM with a validation split of 0.2 and a dropout rate of 0.2. Passes through an word embedding layer first

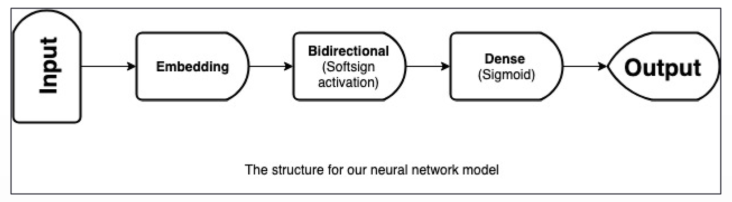

Figure 4. RNN Model

Note: These graphs were obtained from training the model off of the merge set.

Figure 5. RNN Loss

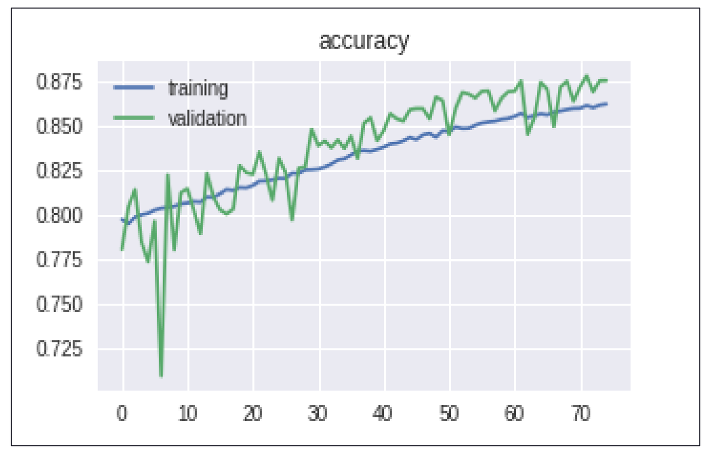

Figure 6. RNN Accuracy

Figure 7. W2V Loss

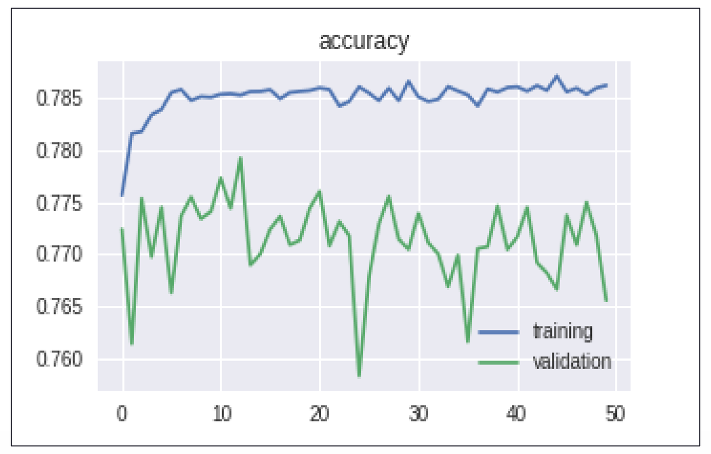

Figure 8. W2V Accuracy

EVALUATION RESULTS

	Accuracy	F1-score	Precision	Recall
SentiWordNet	0.3680	0.5382	0.6601	0.4543
Naïve Bayes (BoW)	0.7605	0.8195	0.8043	0.8353
Naïve Bayes (TF-IDF)	0.7531	0.8303	0.7511	0.9280
Logistic Regression	0.7452	0.8094	0.7915	0.8282
Word2Vec	0.6910	0.7882	0.7055	0.9012
LSTM	0.7469	0.7469	0.7459	0.7469

Table 1. Evaluation Results from merge set

SentiWordNet has the same evaluation results between both sets because of how it was generated

Naïve Bayes was used with Bag of Words and TF-IDF vectors

	Accuracy	F1-score	Precision	Recall
SentiWordNet	0.3680	0.5382	0.6601	0.4543
Naïve Bayes (BoW)	0.6703	0.6864	0.6198	0.7702
Naïve Bayes (TF-IDF)	0.6314	0.6985	0.5655	0.9178
Logistic Regression	0.6603	0.6882	0.6029	0.7704
Word2Vec	0.5674	0.3831	0.4706	0.3248
LSTM	0.6330	0.6331	0.6332	0.6333

Table 2. Evaluation Results from original train and test sets

SentiWordNet has the same evaluation results between both sets because of how it was generated

Naïve Bayes was used with Bag of Words and TF-IDF vectors

CONCLUSIONS

- When trained on the original sets, we were unable to replicate a F1-scores with any of our models comparable to the highest F1-score of 0.75 generated by the NLDS classifiers
  - All of our models except for SentiWordNet and Word2Vec were able to achieve Pos F1-scores on par with most of the NLDS classifiers
  - We noted better F1-scores, higher than those obtained by the NLDS classifiers, when we trained our data off of the merged set. This may indicate that the original test set does not accurately reflect the training set
- Our Naïve Bayes model had the best evaluation results compared to other methods
  - the Naïve Bayes model trained on the TF-IDF vector scored the highest recall score consistent between the merge and original sets
  - However, the model trained on the TF-IDF vector had a significantly lower precision, which may mean that the TF-IDF vectorizer is eliminating too many significant results
- Truth matrices of the LSTM, Naïve Bayes, and Logistic Regression models showed that models had a harder time identifying negative sentiments
  - We concluded that this was because of the unbalanced data set, which had significantly more positive sentiments present
- The LSTM had nearly identical scores for each of the different categories
  - Our group was unable to conclude as to why these scores were so similar
  - We feel that, if trained for more epochs, the LSTM has the potential to have better evaluation results than the Naïve Bayes model
- Future Improvements to our approach
  - Balance the data
  - Train the LSTM for more than 150 epochs
  - Use WordNet-Affect to analyze emotions instead of just sentiments

REFERENCES

Besbes, Ahmed. “Sentiment analysis on Twitter using word2vec and keras.” *Ahmed BESBES – Data Science Portfolio*, 20 Apr. 2017, ahmedbesbes.com/sentiment-analysis-on-twitter-using-word2vec-and-keras.html.

J. Wang, T. Liu, X. Luo, and L. Wang. “An LSTM Approach to Short Text Sentiment Classification with Word Embeddings.” (2018).

L. Reed, J. Wu, S. Oraby, P. Anand, and M. Walker. ”Learning Lexical-Functional Patterns for First-Person Sentiment.” in the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Vancouver, Canada. 2017.