How Will Your Tweet Be Received?

Predicting the Sentiment Polarity of Tweet Replies

NYU Center for Data Science

Project Team: Isidora Filipovic, Alex Herron, Kristin Mullaney, Joseph Schuman

Introduction

Hypothesis

Deep learning models can be trained to predict the aggregate sentiment of a primary tweet's replies, based on the text of the primary tweet.

Background

Predicting the sentiment of tweets has become a very popular task in natural language processing in recent years.

However, substantially less research has focused on predicting the sentiment of a given tweet's replies (which can be used to quantify how a tweet was received).

Relevant work

For this project, we replicated the paper "*How Will Your Tweet Be Received? Predicting the Sentiment Polarity of Tweet Replies*" [1]

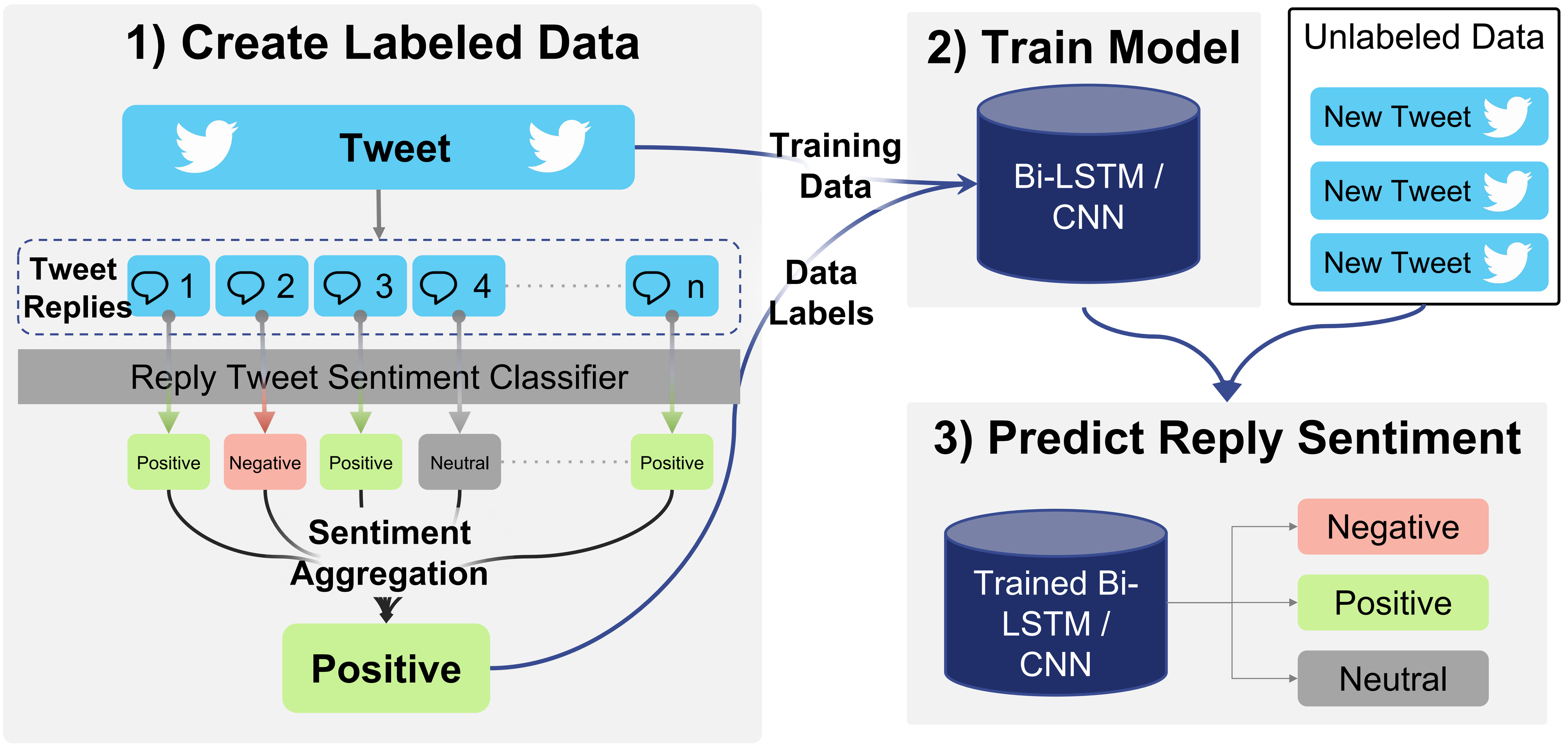
Data Preprocessing

- The data consists of text and metadata of primary tweets (collected via Twitter API using tweet IDs), an aggregate sentiment label, and the sentiment labels of associated first order reply tweets.
- Due to several preprocessing issues in the aggregated sentiment label provided by the original authors (duplicates, missing IDs), we leveraged the reply sentiment labels to create our own aggregated sentiment labels using the same logic as the original authors. A cleaned version of the original and re-processed datasets was also created to address special characters, links, etc. in tweet text. All models were trained on each of the resulting four datasets.
- The table below summarizes class balance and size of the primary tweets in the re-processed and original training sets.
- The aggregation method assigns positive and negative labels to tweets that possess overwhelmingly positive and negative responses, respectively.

  The neutral label is assigned to those tweets that either have a significant amount of neutral reply tweet labels (over 85%) or garnered relatively even positive and negative responses.
- The test set for our Bi-LSTM and CNN classifiers was a manually annotated "gold label" dataset

Sentiment Label	Re-processed	Original
Positive	12,149	5,237
Neutral	12,149	10,659
Negative	8,258	7,181
Total	**26,655**	**23,077**

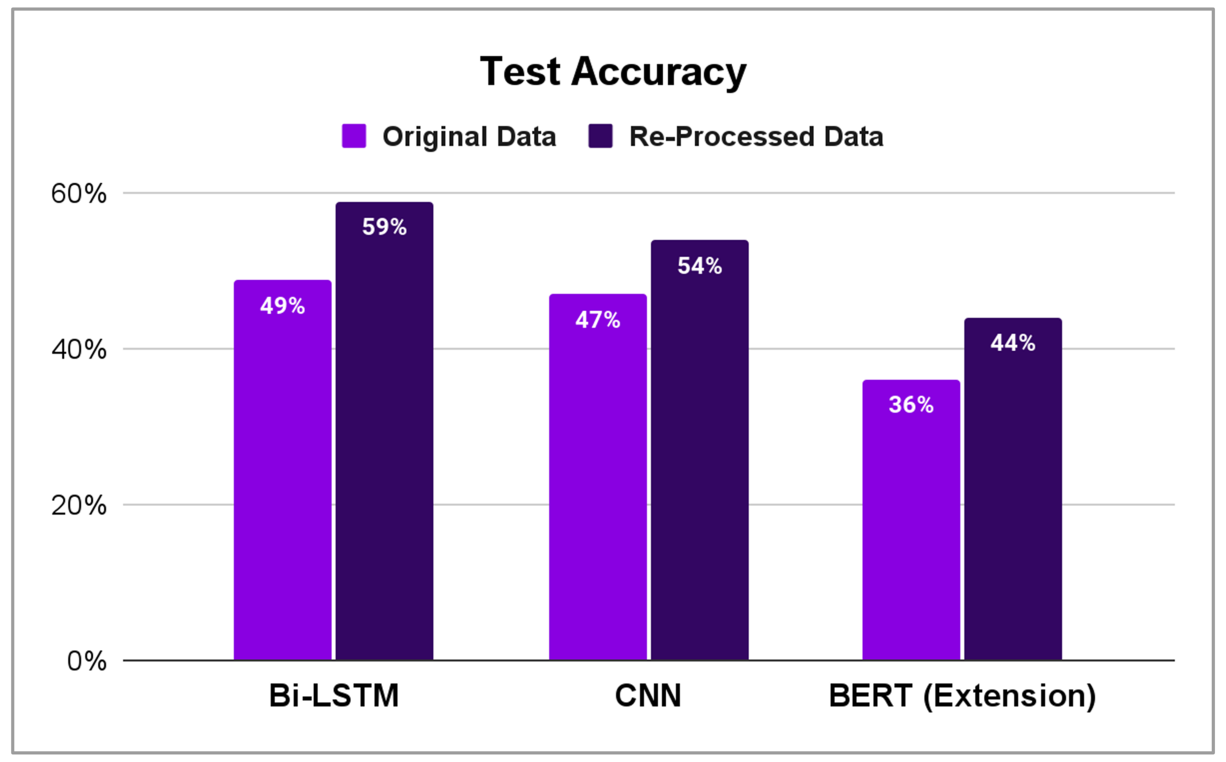
Data Pipeline



Results and Extensions

Results

- Trained Bi-LSTM and CNN on each of our four datasets
- The below chart highlights results of models trained on the original and re-processed data (which got better results than the original data provided)
- Our best performing model (a Bi-LSTM, with accuracy of 59%) was out-performed by the original paper's best model (a CNN), which reported a peak accuracy of 61%



Extension 1: Sentiment Analysis

- The test (manually-labeled) tends to have more positive sentiment than training data (see table below)
- Distribution of primary tweet sentiment is different from aggregate reply sentiment -> primary tweet sentiment doesn't always match reply sentiment

Reply Sentiment	Train	Test
Negative	45.6%	38.7%
Neutral	31.0%	31.9%
Positive	23.4%	29.4%

Extension 2: Exploratory Data Analysis

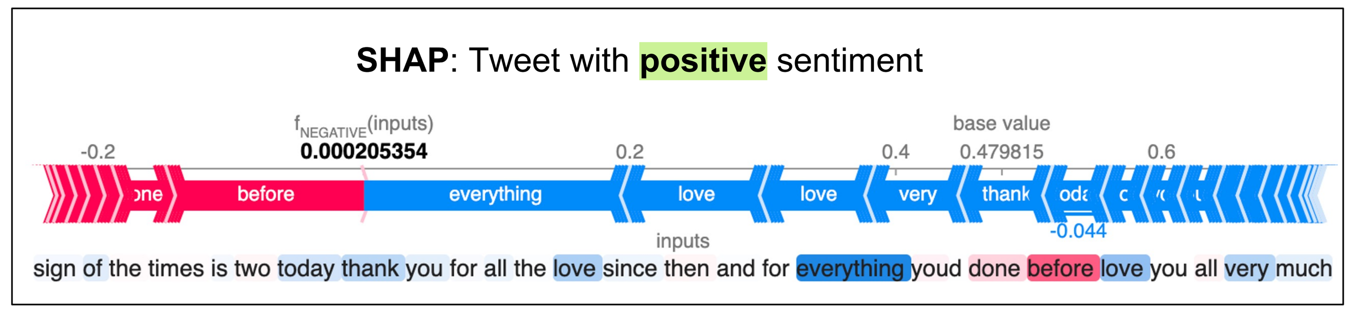
- Disproportionately negative tweet sentiment on Saturdays
- Average number of replies for negative tweets was double that of positive/neutral tweets

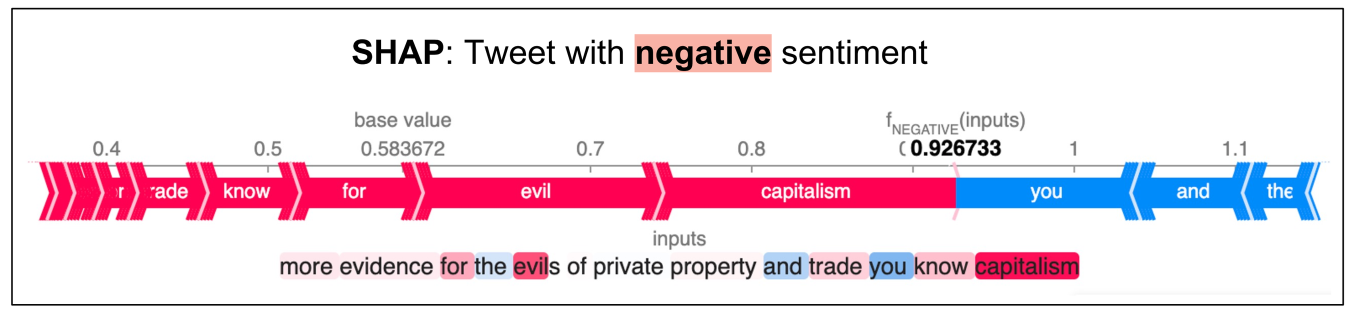
Extension 3: Pretrained Model Using BERT

- Limited to only 10 epochs for timing's sake
- Results on left. Further training likely to increase accuracy on both original and re-processed data.

Extension 4: SHAP

- SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of machine learning models [2]
- SHAP can provide insight as to the specific words or tokens that contribute to a positive or negative sentiment
- Below a word like "evil" contributes to negative sentiment
- However, less intuitively, "capitalism" also contributes to negative sentiment
- Words like "love" strongly contribute to positive sentiment
- Further analysis can identify the words that contribute most frequently to misclassification





Conclusions

- **Deep learning models can be trained to predict the aggregate sentiment of a primary tweet's replies, using the tweet's text as input**
- Our Bi-LSTM trained using our re-processed data obtained the best results, with an accuracy of 59%
- Preprocessing is absolutely crucial for working with Twitter data, particularly in the context of labeling data in an automated manner

References

[1] Arasteh, S.T. et al. (2021) "How will your tweet be received? predicting the sentiment polarity of tweet replies," 2021 IEEE 15th International Conference on Semantic Computing (ICSC) [Preprint]. Available at: https://doi.org/10.1109/icsc50631.2021.00068.

[2]BERT. Available at: https://huggingface.co/docs/transformers/model_doc/bert (Accessed: October 20, 2022).

[3] Welcome to the SHAP documentation - SHAP latest documentation. Available at: https://shap.readthedocs.io/en/latest/index.html (Accessed: October 20, 2022).

[4] PyTorch documentation - PyTorch 1.12 documentation. Available at: https://pytorch.org/docs/1.12/ (Accessed: October 20, 2022).

[5] How to convert tweet id to the actual tweet and associated metadata (no date) twittercommunity.com. Available at: https://twittercommunity.com/t/how-to-convert-tweet-id-to-the-actual-tweet-and-associated-metadata/157648.

[6] DocNow (no date) DocNow/TWARC: A command line tool (and Python Library) for archiving twitter JSON, GitHub. Available at: https://github.com/DocNow/twarc (Accessed: November 16, 2022).

[7] Mining replies to tweets: A walkthrough - towards data science (no date). Available at: https://towardsdatascience.com/mining-replies-to-tweets-a-walkthrough-9a936602c4d6 (Accessed: November 17, 2022).

[8] GETOLDTWEETS3 (no date) PyPI. Available at: https://pypi.org/project/GetOldTweets3/ (Accessed: November 17, 2022).

[9] chayan8 (2020) *Sentiment analysis using Bert: Pytorch*, *Kaggle*. Kaggle. Available at: https://www.kaggle.com/code/chayan8/sentiment-analysis-using-bert-pytorch (Accessed: December 7, 2022).