COVID-19 Twitter Sentiment Analysis Over Time

Shihao Niu
asher.niu@cs.ucla.edu

Thomas Pan
thomascpan@g.ucla.edu

Travis Tu
ttu18@g.ucla.edu

Abstract

In this paper, we describe our attempts to produce sentiment classifiers and to analyze the sentiments of COVID-19 related tweets. We train two shallow machine learning model using support vector machines (SVM) and Logistic Regression, and two deep learning NLP models based on Convolutional Neural Network (CNN) and Long short-term memory (LSTM), on a generic data set of tweets. We evaluate the performance of these models on the generic tweets, use these pre-trained models to predict the sentiment of COVID-19 tweet, and analyze how these COVID-19 sentiments vary over the duration of the pandemic by month. We hypothesize that due to the explosion of tweets and the non-uniform distribution of data, sentiments regarding COVID-19 have become more negative over time. However, even our best models (CNN) need fine-tuning on labeled COVID-19 tweets and pre-training on neutral sentiment to perform accurate temporal sentiment analysis of COVID-19 tweets.

1 Introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease that was first identified in Wuhan, China in December 2019. COVID-19 has since rapidly spread, causing a global pandemic. As of May 2020, there have been over 5.21 million COVID-19 cases which have resulted in 338 thousand deaths. In the United States alone, there are currently 1.3 million cases and 96 thousand deaths. The rapid proliferation of cases worldwide has resulted in lock-downs in many countries, straining the global economy and people's livelihoods.

As one of the most disastrous events in modern history, it is crucial to analyze the effects that COVID-19 has on people. We explore the sentiments of COVID-19 tweets because extracting opinions, feelings, and emotions from a body of text can be a valuable gauge for everyone from investors evaluating stock futures to politicians trying to gain an edge in an upcoming election, and perhaps most importantly, healthcare professional monitoring the deleterious effects of a crisis on mental health. We choose Twitter as our medium because in a digitized world, people often express themselves via social media, and for the hundreds of millions of people worldwide experiencing lockdown, social media has become one of the few ways to communicate.

In the present work, we train a SVM model and a logistic regression model, both of which proved insufficient for our ultimate goals. Thus, we pretrain two sentiment classifiers, CNN and LSTM, on a corpus of generic Twitter data, and utilize them to examine people's sentiment towards COVID-19 by analyzing their tweets and observing the variation of positive versus negative sentiments over time. Our best model, the CNN model, shows promise with fine-tuning on manually labelled COVID-19 tweets, but our results also suggest the need for an additional neutral sentiment or finer-grained sentiment analysis.

As a note to Professor Chang, this is our first natural language processing course and our first time working with deep learning models such as LSTMs and CNNs. Thus, we focused on techniques that we have learned in class or have encountered in our literature survey.

2 Related Works

2.1 Twitter and COVID-19

In the social media age, Twitter has become a direct source of an unprecedented amount of content ranging from minor status updates to the political exploits of the United States government. Most of the current work on COVID-19 related tweets have been focused on gathering information or studying the propagation of misinformation. A multilingual COVID-19 tweet corpus dating back to January 22, 2020 has been made publicly available and continues to be updated (Chen et al., 2020). Researchers have also studied the themes and sources of COVID-19 tweets and their connection to both reliable and unreliable Internet sources (Singh et al., 2020). As COVID-19 has stirred up global discourse on social media, models of the spread of COVID-19 rumors on social media with epidemiological models have been made (Cinelli et al., 2020).(Barkur and Vibha, 2020) have performed sentiment analysis of COVID-19 tweets, but only to gauge the Indian population's reaction to a lockdown. Recently, (Abd-Alrazaq et al., 2020) analyzed general themes of tweets such as increased racism towards East Asians, but did not study the sentiment of individual tweets over the course of the COVID-19 crisis. To the authors' best knowledge, this is the first work studying the long-term time distribution of COVID-19 tweet sentiments.

2.2 Twitter Sentiment Classification

Twitter sentiment classification has become an increasingly popular NLP research topic. Some groups opted to take a lexical approach to this problem. By using dictionaries of known sentiment-bearing words and phrases annotated with polarity and strength of each word, they created classifiers that only exploit direct sentiment indicators (Thelwall et al., 2012). (Taboada et al., 2011) further incorporated negation and intensification to extract sentiments from text. Twitter-specific sentiment lexicons were often built using distant supervision where emoticons can serve as noisy labels of a tweet's sentiment (Go et al., 2009). These emoticons can be used to simply smooth the language model (Liu et al., 2012) or used to capture relations between phrases and sentiment seeds such as emoticons and hashtags (Mohammad et al., 2013).

The lexicon-based approaches soon were combined with machine learning based approaches where the lexicon-based sentiment score became a feature in a machine learning classifier such as SVM or Naïve Bayes (Kolchyna et al., 2015) or used to label the training data of such shallow machine learning classifiers (Zhang et al., 2011). Domain transferable Twitter specific lexicons have also created features that proved useful in getting accurate Twitter sentiment analysis results using DAN2 and SVM (Ghiassi and Lee, 2018). However, these approaches generally require laborious feature engineering that requires expert domain-specific knowledge and still may result in redundant and missing features.

Deep neural networks, when provided sufficient training data, can automatically learn features useful for both general and Twitter sentiment classification. (Kim, 2014) used CNNs to learn task-specific vectors through fine-tuning and combined them with static vectors to achieve state-of-the-art results at the time for sentiment analysis. To analyze short text sentiments with limited contextual information such as tweets, (Dos Santos and Gatti, 2014) used a CNN that extracted character and sentence level information from two corpora from different domains.

Recurrent neural networks (RNNs) have also been used for Twitter sentiment classification, though simpler versions are being replaced by bidirectional LSTMs to address the gradient vanishing problem. For example, (Yang et al., 2017) used an attention-based bidirectional LSTM approach to improve target-dependent sentiment classification for tweets and outperformed (Dong et al., 2014)'s adaptive RNN that selectively propagated word sentiments depending on the syntactic relationships between the word and the target. LSTM's use of gates and memory cells allowed models using them to outperform vanilla RNNs on more complex sentiment expressions by simulating the interactions of longer sequence of words during tweet composition to predict tweet polarity (Wang et al., 2015). LSTMs can be combined with other techniques as well. For example, (Chen et al., 2018) trained an attention-based LSTM by initializing weights using the Vader rule-based sentiment analysis algorithm, fine-tuning on labelled data, and incorporating positive and negative emoji embeddings.

Some models now even combine multiple deep learning approaches. (Sosa, 2017) combined LSTM and CNN models to improve performance compared to just LSTM or just CNN models and also fine-tuned models initialized with pre-trained GloVe word embeddings. However, this was limited by misspellings, word abbreviations, internet slang that are common in tweets, but are not in GloVe word embeddings. The best performing model from SemEval-2017's Twitter sentiment analysis task used an ensemble of ten CNNs and ten LSTMs trained via transfer learning from pre-trained word embeddings from Word2Vec, GloVe, and FastText and fine-tuning word embeddings by using emoticons to weakly label sentiment polarity in a distant training phase. (Cliche, 2017). Other work connects deep CNNs that exploit character level information for word-level embedding to a bidirectional LSTM to produce a sentence-wide feature representation (Nguyen and Nguyen, 2017).

3 Methodology

We train our sentiment analysis classifier on a generic Twitter data set from Stanford called Sentiment140 (Go et al., 2009), because there is no labeled COVID-19 twitter data set currently available. Given the scope and the time of the project, it is also infeasible to manually label enough COVID-19 Twitter data ourselves to train on from scratch. The data set consists of 1.6 million tweets and is well-balanced between positive and negative sentiment. The data is split 80/10/10 into training, validation, and test, respectively.

Afterwards, we use the COVID-19 specific Twitter data set from (Chen et al., 2020). The data set consists of approximately 140 million COVID-19 related tweets gathered from Twitter's API from user accounts by searching for related keywords such as #COVID19. Tweets were gathered starting from late January to late May.

Both sets of tweets are pre-processed by using the Beautiful Soup library and the lxml tool kit to parse and remove HTML formatting. If present, UTF-8 byte order marks are removed. The "@" and "#" symbols and URLs are removed using regex pattern matching. All alphabetical characters are converted to lower case. We use the WordPunctTokenizer from the nltk library to tokenize each tweet. However, this tokenizer removes the apostrophe and "t" from negation contractions. For example, "can't" would become "can" after tokenization, which would likely flip the sentiment of a tweet. To fix this, we use a dictionary of common negation contractions and regex pattern matching to convert them to the longer form of the contraction prior to tokenization (Kim, 2017a). The tokens are concatenated back together, separated by spaces, to create the final pre-processed tweet. Several tweets that contained only URLs became null entries in the data set and are consequently filtered out. Even though emoticons can serve as a useful markers for tweet polarity, we remove them in this model.

The best model is further fine-tuned using a small COVID-19 related Twitter data set of 308 positive sentiment tweets and 308 negative sentiment tweets that we manually labelled. We decide disagreements on the sentiment of a tweet by majority vote. The data is split 80/10/10 into training, validation, and test, respectively. The best fine-tuned model in terms of accuracy and F1 score is saved.

Lastly, we apply our generic-Twitter data set pre-trained classifiers and the fine-tuned classifier to the COVID-19 data set and explore the sentiments behind. We perform sentiment analysis on a random sample of approximately 350 thousand tweets. We filter out tweets of more than 140 characters because the generic Twitter data set we pre-trained on was from 2014, before Twitter increased the tweet character limit to 280 characters in 2018. We wish to explore the distribution of different sentiments over the different months, and see if our hypothesis that the change of the scope of the outbreak and the Twitter data is correlated to the change of the sentiment distributions.

4 Proposed Experiments

4.1 SVM and Logistic Regression

We initially experiment with training SVM and logistic regression shallow machine learning models on the generic Twitter data set. Specifically, we pre-process the tweets and use GloVe to generate the word embedding vector for each word in the tweets. To handle out-of-vocabulary (OOV) words, we elect to use the mean of the all embeddings in the vocabulary, instead of random or all zero embeddings. Then, we use the average of all the embeddings within a tweet to represent the embedding of said tweet. These embeddings will be used as the features of the tweet we will train our classifiers on.

4.2 LSTM

LSTMs are special RNNs. Unlike traditional neural networks which are unable to handle sequential information, RNN includes recurrent connections inside its network. This looping constraint allows it to capture sequential information. However, RNNs experience the vanishing gradient problem which causes gradients to become too small, essentially forgetting old context. LSTM are specifically designed to overcome this problem by remembering states. LSTMs are composed of cells and gates. Cells remember states over time and the gates control what information is added and removed. For our implementation, we use an embedding size of 32, a hidden state size of 64, two recurrent layers, a dropout of 0.3 at the embedding layer, and a dropout of 50% at the LSTM layers. Training is performed over four epochs at a batch size of 32. Loss is calculated via binary cross-entropy, and optimization is performed using the Adam optimizer with performance evaluated by F1 scores and accuracy (Agrawal, 2019).

4.3 Convolutional Neural Network

We use the Word2Vec API from the genism library with the skip-gram option and a window size of five to create Word2Vec word embedding vectors with 200 dimensions from the training data. By learning to predict the context given a word, the model will be able to learn to understand even infrequently encountered words and the larger window should help capture more topic and domain information, both of which should be useful when applying the model to COVID-19 related tweets.

The tweets are transformed into a sequence of integers by the text_to_sequences function from the Keras library and the pre-trained Word2Vec word embeddings are concatenated into an embedding matrix after being padded so all the tweets had the same matrix dimension (Kim, 2017b). After the embedding layer, three convolutional layers with 200 filtering matrices and filter sizes of three, four, or five were are applied to learn n-grams in the word embedding space. A max pooling operation allows the CNN to effectively extract the most important n-grams found by the convolutions and pass them through a dense fully connected hidden layer and finally through a layer with a sigmoid function to perform binary classification. To reduce over-fitting, dropout layers with 50% dropout probability are added both before and after the final fully-connected hidden layer. Loss is calculated as binary cross-entropy. Optimization is performed with the Adam optimizer. We train over five epochs with a batch size of 32 and evaluate performance using F1 scores and accuracy.

4.4 Fine-Tuning

After evaluating all the previous models on the generic Twitter data set, the best model is selected to be fine-tuned with our manually labelled COVID-19 tweets. The CNN is the chosen model so fine-tuning requires to removal of the final binary classification layer from the original model and the unfreezing the weights in all the previous layers. A fully connected hidden layer of 256 nodes and applying the ReLU activation function is added followed by another 50% dropout layer to prevent over-fitting. Finally, a layer with a sigmoid function to perform binary classification is added. The fine-tuning is performed with batch sizes of eight and over ten epochs because the labelled COVID-19 data set is very small. The best fine-tuned model in terms of accuracy and F1 score was saved and used to get the time distribution.

4.5 Time Distribution

Lastly, to answer our questions raised in our hypothesis, we use our trained classifier to classify our COVID-19 data set. We feed each classifier with data from the five months of the COVID-19 pandemic so far to get a distribution of positive and negative sentiments.

5 Results

5.1 SVM and Logistic Regression

We are unable to achieve satisfying results with our SVM and Logistic Regression models. One of the major shortcomings is that during training, the classifier needs to constantly look up the word embeddings from Word2Vec/GloVe to interpret the sentence embeddings. This step turns out to be a major bottleneck in both our training and test stage. In addition, this high overhead manifested again during prediction on the COVID data set, where we have to look up the embeddings for the input as well. Without storing the data in a proper database that allows high speed lookup, we ruled these models out as infeasible. When the code was run on the much smaller SemEval2017 Task 4 data set, the classifiers were not able to perform much better than random guessing.

5.2 LSTM

When trained on the generic Twitter data set, our LSTM is able to achieve a test accuracy of 75.3% and a F1 score of 0.596 as shown in Table 1. Having pre-trained our LSTM model on the generic Twitter data set, we apply our model to the COVID-19 related tweets. Our LSTM model finds that approximately 52.5% of all tweets carry a positive sentiment. In addition, sentiment shows little variation over the five months of the COVID-19 crisis as shown in Figure 1. below.

	LSTM	CNN
F1 Score	0.596	0.812
Accuracy	75.3%	82.1%

Table 1: Model metrics on generic Twitter data set

	Pre-trained CNN	CNN Fine-tuned
F1 Score	0.46	0.83
Accuracy	46.77%	75.81%

Table 2: Model metrics on manually labeled COVID-19 Twitter data

5.3 CNN and Fine-tuning

After training the CNN on the generic Twitter data set, the CNN achieves a test accuracy of 82.1% and a F1 score of 0.812 with a precision of 0.86 and a recall of 0.77 as shown in Table 1. Combined with the confusion matrix, it appears that the CNN tends predict many more false negatives than false positives. When applied to the unlabelled COVID-19 Twitter data set, the CNN determines that approximately 40.5% of the tweets have a positive sentiment, which is consistent with the model predicting more false negatives than false positives. Like the LSTM, the CNN shows little variation over the course of the pandemic as can be seen in Figure 1 below.

When the pre-trained CNN is tested on the test set of the manually labelled COVID-19 tweets, it achieves an accuracy of 46.77% and a F1 score of 0.46 with a precision of 0.70 and a recall of 0.34. This is not unexpected as the COVID-19 tweets are likely to be a different distribution than the generic Twitter data set. After fine-tuning the CNN on the small set of manually labelled COVID-19 tweet, the results improve dramatically when tested on the manually labelled COVID-19 test data set; the best fine-tuned model has an accuracy of 75.81% with a F1 score of 0.83 as shown in Table 2 below. Interestingly, the precision is 0.78 and the recall is 0.52. When combined with the confusion matrix, it seems that fine-tuning has now caused the CNN to over-predict false positives instead of false negatives. The ratio of positive to negative sentiment generally has insignificant fluctuation, but there is now a small peak in February indicating more positive sentiment in the second month of the pandemic.

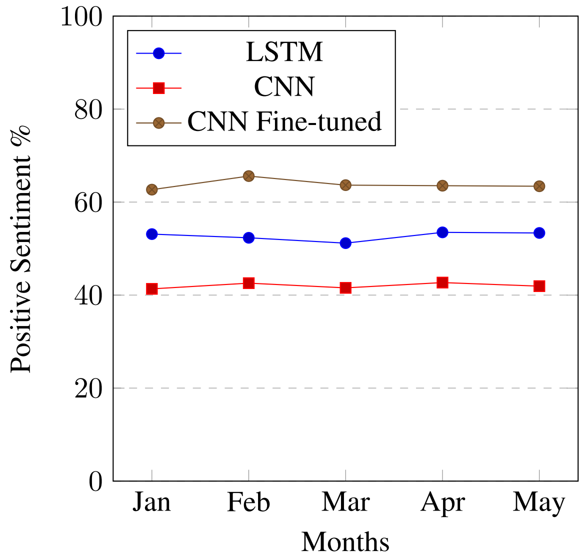

Figure 1: Percent of COVID-19 tweets with positive sentiment

6 Discussion

We manually examined some of the tweets to get a better understanding of our mixed and unintuitive results. In both the CNN and LSTM models we consistently find misclassified tweets that we interpret as slightly positive or slightly negative. The tweet below is predicted to be positive by the LSTM, when it appears to be slightly negative.

> RT @RepMichaelWaltz: From the beginning, #China's Communist Party hasn't been honest about #COVID19 with the @WHO, @CDCGov or the world.

We also find many tweets from journalistic sources such as the New York Times or the Wall Street Journal that read as matter-of-fact news headlines that should classified as neutral sentiment. The tweet below is classified as negative by the pre-trained CNN, but it can be interpreted as just factual reporting.

> RT @WSJ: People infected with coronavirus may have no symptoms at all. That's why the virus is so hard to stop. https://t.co/55VdAHpvbH

The following tweet is also classified as negative by the LSTM, but appears to carry a neutral sentiment.

> RT @FrankCaliendo: SOUND ON! Jim Nantz and Tony Romo Do Play By Play of the Coronavirus Expert's Press Conference. https://t.co/u28qOzqM2v...

The small spike in positive sentiment in February might be explained by the emergence of the first cases of COVID-19 in the United States. We do not believe it is an actual increase in positive sentiment related to COVID-19 as we judged far more negative sentiment than positive sentiment tweets when manually labelling our COVID-19 Twitter data set. In the early days of the emergence of the disease on United States soil generated much more interest from reporters, leading to a spike in neutral sentiment headline posts that are more likely to be classified as false positive by the fine-tuned CNN, such as the tweet show below. Perhaps COVID-19 news fatigue set in or other events such as the Black Lives Matter movement came to dominate the headlines in subsequent month.

> RT @BNODesk: Japanese researchers estimate that at least half of new infections of coronavirus occur while the first patient is not showing symptoms.

By spot checking examples such as the ones shown above, we notice that many tweets have neutral sentiment or have much more subtle sentiment than the tweets in our generic Twitter training set. As such, our future models may benefit from pre-training on a data set that includes a neutral sentiment and possibly even finer-grained sentiment classification labels. Because of this observation, we exclude neutral sentiment tweets when creating the manually labeled COVID-19 Twitter data set, but keep tweets with subtle sentiments.

The fine-tuning, even though it is only on a few hundred COVID-19 related tweets, probably helps the pre-trained CNN better classify the sentiments of tweets because the Stanford Sentiment140 data set is likely missing Internet lexicon that did not exist in 2014. Terms such as "boomer" and "fake news" did not exist in 2014 and terms such as "Donald Trump" and "coronavirus" were much less frequent or in different contexts. Thus, in addition to adding neutral and finer-grained sentiment classification, we would like to expand our manually labeled COVID-19 Twitter data set to several thousand tweets in the future to further refine our models.

In addition, we notice that all our models struggle to correct classify the sentiment of sarcastic tweets. For instance, the tweet below was classified a positive even though its sarcastic tone makes it a negative sentiment.

> RT @tbonnita: Yeah, I AM TOTALLY ready for another lockdown weekend Me,SUNDAY at 3pm. #FrontRoomChatWithSirAludah reloaded as @DjAludah

In the future, we can better address sarcasm and improve our model in general by incorporating binarized emoticons and hashtags to use distant supervision to weakly label tweets ahead of time in our pre-trained models as was done in (Felbo et al., 2017). Furthermore, we can use a bidirectional language model such as BERT to pre-train our deep learning models by conditioning on both left and right contexts of tweets. For this sentiment classification setting, we input a single tweet and the special CLS token into the BERT model and take the output of the transformer for the CLS word embedding as the representation of the tweet (Devlin et al., 2018). This can then be fed directly into a softmax classification layer or used as input into our CNN model. Finally, we can combine all our models into an ensemble model and add different CNNs with various filter sizes to extract different n-grams in the word embedding space as shown by (Cliche, 2017).

7 Conclusion

In this paper, we presented several models to perform binary sentiment classification of COVID-19 related tweets and applied them to analyze COVID-19 related sentiment over the five months of the pandemic. Because no sentiment-labelled COVID-19 Twitter data set exists at this time, we pre-trained our models on a generic twitter sentiment classification data set and fine-tuned our CNN model on a small set of manually labelled COVID-19 tweets. Though we currently find little temporal variation in binary sentiment, our approach showed promise in terms of obtaining more robust and trustworthy results in the future with more data and by pre-training the models to include neutral sentiments and finer-grained sentiment classification.

References

Alaa Abd-Alrazaq, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. 2020. Top concerns of tweeters during the covid-19 pandemic: infoveillance study. *Journal of medical Internet research*, 22(4):e19016.

Samarth Agrawal. 2019. Sentiment analysis using lstm (step-by-step tutorial).

Gopalkrishna Barkur and Giridhar B Kamath Vibha. 2020. Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india. *Asian journal of psychiatry*.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Covid-19: The first public coronavirus twitter dataset. *ArXiv*, abs/2003.07372.

Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. 2018. Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 117–125.

Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. The covid-19 social media infodemic. *arXiv preprint arXiv:2003.05004*.

Mathieu Cliche. 2017. Bb_twtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms. *arXiv preprint arXiv:1704.06125*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, pages 49–54.

Cicero Dos Santos and Maira Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 69–78.

Dimitrios Effrosynidis, Symeon Symeonidis, and Avi Arampatzis. 2017. A comparison of pre-processing techniques for twitter sentiment analysis.

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*.

Manoochehr Ghiassi and S Lee. 2018. A domain transferable lexicon set for twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, 106:197–216.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.

Ricky Kim. 2017a. Another twitter sentiment analysis with python — part 1.

Ricky Kim. 2017b. Another twitter sentiment analysis with python — part 11 (cnn + word2vec).

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Olga Kolchyna, Tharsis TP Souza, Philip Treleaven, and Tomaso Aste. 2015. Twitter sentiment analysis: Lexicon method, machine learning method and their combination. *arXiv preprint arXiv:1507.00955*.

Kun-Lin Liu, Wu-Jun Li, and Minyi Guo. 2012. Emoticon smoothed language models for twitter sentiment analysis. In *Twenty-sixth aAAI conference on artificial intelligence*.

Saif M. Mohammad, Svetlana Kiritchenko, and Xiao-Dan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *SemEval@NAACL-HLT*.

Huy Nguyen and Minh-Le Nguyen. 2017. A deep neural architecture for sentence-level sentiment classification in twitter social networking. In *International Conference of the Pacific Association for Computational Linguistics*, pages 15–27. Springer.

Lisa Singh, Shweta Bansal, Leticia Bode, Ceren Budak, Guangqing Chi, Kornraphop Kawintiranon, Colton Padden, Rebecca Vanarsdall, Emily Vraga, and Yanchen Wang. 2020. A first look at covid-19 information and misinformation sharing on twitter. *arXiv preprint arXiv:2003.13907*.

Pedro M Sosa. 2017. Twitter sentiment analysis using combined lstm-cnn models. *Eprint Arxiv*.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.*, 63(1):163–173.

Xin Wang, Yuanchao Liu, Chengjie Sun, Baoxun Wang, and Xiaolong Wang. 2015. Predicting polarities of tweets by composing word embeddings with long short-term memory. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1343–1353, Beijing, China. Association for Computational Linguistics.

Min Yang, Wenting Tu, Jingxuan Wang, Fei Xu, and Xiaojun Chen. 2017. Attention based lstm for target dependent sentiment classification. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Lei Zhang, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and Bing Liu. 2011. Combining lexicon-based and learning-based methods for twitter sentiment analysis.