COVID-19 SENTIMENT ANALYSIS

COLX 585 001 2020W Trends in Computational Linguistics Final Paper

Mentor: Chiyu Zhang

Team-3: Yundong Yao, Alex Chen, Quang Vu, Lisa Liu

1. INTRODUCTION

Since the end of 2019, COVID-19 broke out gradually country by country throughout the world, almost no place on earth could stay out of this epidemic. Human got different feelings: Some feel frightened, angry, and desperate, while the others feel that it is just a bad cold, and sooner or later, human beings will overcome the virus. From the course of Trends in Computational Linguistics, we got the knowledge of the most advanced neural network model and natural language processing method. We would like to practice them to the Covid-19 related social media data to do sentiment analysis, so that to find out the best model for this kind of application. At the same time, we are also curious about in the same world, do people in different country got the same feeling towards the pandemic, and the role of time and space play.

2. PREVIOUS WORK

There are many researchers working on sentiment analysis on social media data, including the twitter platform, which give us lots of inspirations:

H. Manguri, K., N. Ramadhan, R. and R. Mohammed Amin, P. [1], published “Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks”, which is based on the analysis of tweets data scraping from Twitter social media on two specified hashtag keywords, (“COVID-19, coronavirus”). The date of searching data is seven days from 09-04-2020 to 15-04-2020, 530232 tweets were collected. In this paper, "Emotional Guidance Scale” for polarity evaluation is setup for the analysis. The study shows that people's reactions vary day to day from posting their feelings on social media specifically Twitter.

Mathieu Clich’és [2], described the Twitter sentiment classifier using Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) networks. They leveraged a large amount of unlabeled data to pre-train word embeddings. They then use a subset of the unlabeled data to fine tune the embeddings using distant supervision. To boost performances, they ensembled several CNNs and LSTMs together. They got 0.685 average F1 score.

3. DATA AND PRE-PROCESSING

3.1 Training dataset

The dataset we used to train our model is the SemEval-2017 dataset. This dataset contained 43k tweets for training set, 2k tweets for validation set and 2k tweets for test set. All the tweets had labels by three polarities: positive, neutral and negative.

Table 3.1 The sentiment distribution in the training dataset

	positive	**neutral**	**negative**	**total**
train	17,849	20,673	7,093	45,615
dve	869	819	312	2,000
test	5,937	2,375	3,972	12,284

3.2 Inference dataset

We collected a 675k tweets subset of MegaCov[4] dataset which is a billion-scale dataset from Twitter for studying COVID-19 created by UBC Deep Learning & NLP Lab. All the tweets were written in English from January to May of 2020. Our dataset consisted data posted from 6 countries.

Table3.2 below provides the distribution of our dataset.

	January	**February**	**March**	**April**	**May**	**Total**
US	75K	117K	225K	83K	32K	**536K**
UK	18K	20K	30K	24K	9K	**104K**
Mexico	2K	0K	0K	7K	2K	**12K**
Canada	0.9K	3K	1K	1K	0.5K	**7K**
France	3K	0.9K	6K	0K	2K	**13K**
New Zealand	0.2K	0.2K	0.5K	0.2K	0.1K	**1K**
Total	**100K**	**142K**	**264K**	**117K**	**47K**	**675K**

3.3 Data preprocessing

Since we were dealing with tweets in this project, we needed to do specific tweet text cleaning along with normal text pre-processing. A tweet may contain:

- User ID
- URL

To clean the tweet, we preprocessed the dataset by:

- Replaced all users’ IDs in tweet text to "@user".
- Replaced all URL in tweet text to "<url>".

4. MODEL TRAINING

4.1 Model Selection

To find the best model on our sentiment analysis task, we have implemented four different models which might be good at classification, they are: Linear SVC, CNN, BERT (large) BERTweet. And we tuned the hyper parameters on each model to try to get the best performance on sentiment classification.

Table 4.1 Hyper parameter tunings on classifiers

Model	Hyper parameter tuning
LinearSVC	N-gram(1gram-4grams)
CNN	epochs (3-100), learning rate (0.001-0.1), num of layers (2,3,4), drop out (0.2-0.7)
BERT large	Epochs (1-10), warmup proportion (0.1-0.2), drop out (2e-5 - 2e-3)
BERTweet	Epochs (1-5), learning rate (2e-5 - 2e-3), maximum length (30 – 128)

4.2 Model Training

Given the three-class classification task (positive vs. negative vs. neutral), we trained four separate classifiers. Our model using BERTweet embeddings got the highest F1 score in both validation and test set.

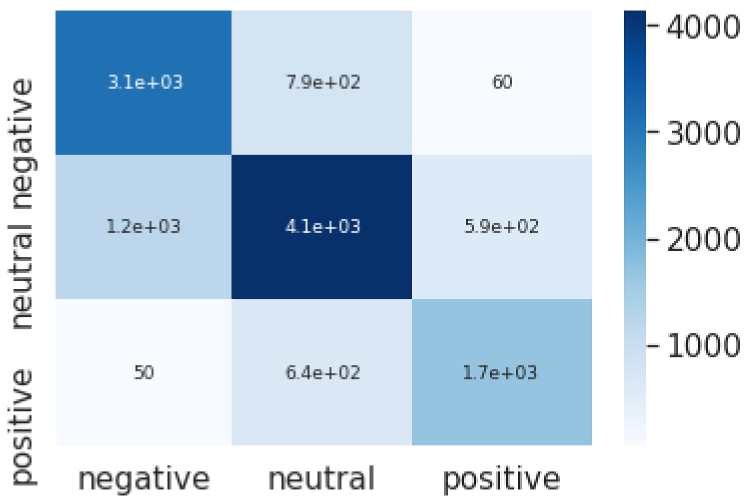
Table 4.2 Accuracy and f1 score of classifiers on validation and test dataset.

Model	Accuracy (Dev)	F1 (Dev)	Accuracy (Test)	F1 (Test)
Dummy Majority (baseline)	0.43	0.43	0.48	0.48
LinearSVC	0.48	0.47	0.43	0.42
CNN	0.51	0.50	0.49	0.46
BERT base	0.63	0.62	0.61	0.61
BERTweet	0.75	0.74	0.73	0.73

The model hyperparameter that we used with BERTweet embeddings are:

- epochs = 3
- learning rate = 2e-5
- max_grad_norm = 1.0
- warmup_proportion = 0.1
- number of hidden states = 768
- drop_out = 0.1

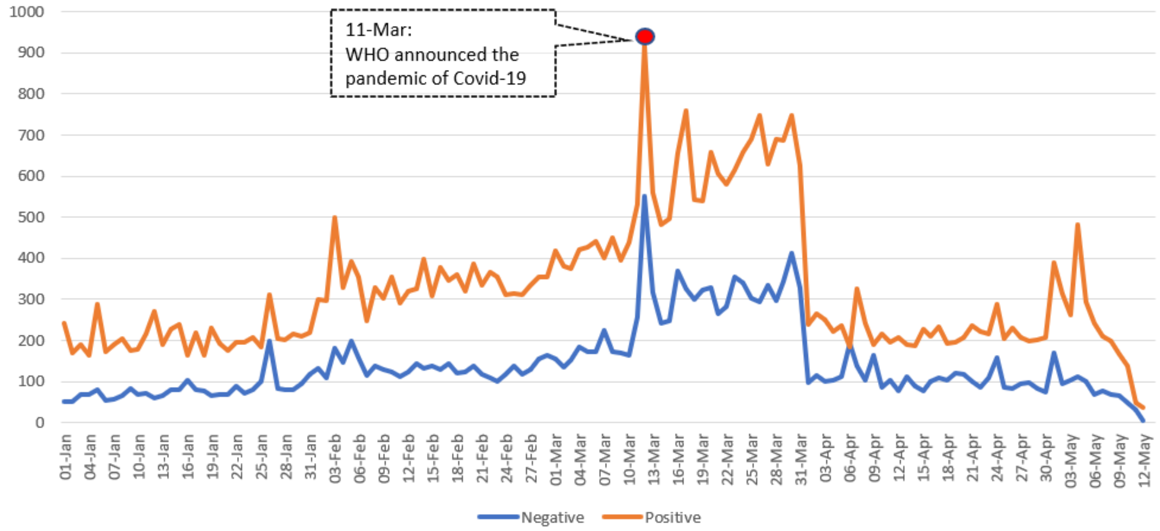
Table 4.3 Confusion matrix on test dataset.



5. FINDINGS

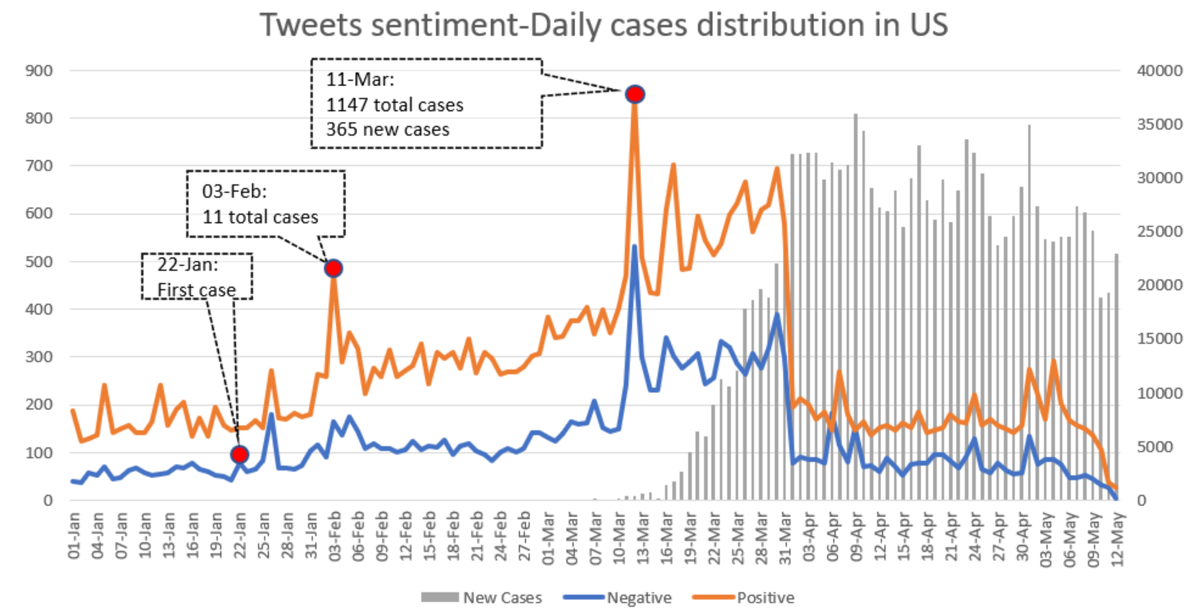
Most of tweets with polarity were in March and the Pollyanna principle were applied with most of tweets are positive. This figure below [5.1] showed the sentiments of our dataset (filtered out neutral tweets). Positive sentiments seemed to prevail most of the time, which could be interpreted as a manifestation of the Pollyanna principle.

Figure 5.1 The number of sentiment tweets changes by time



Also, we found that the sentiment lines in UK [Figure 5.2] is less sensitive to the Covid-19 new cases than the sentiment lines in US [Figure 5.3].

Figure 5.2 Tweets Sentiment-Daily cases distribution in US



When we added new cases by time in the same plot, there seemed to be certain correlation among these time series. The plot above was in regard to the United States.

Figure 5.3 Tweets Sentiment-Daily cases distribution in UK

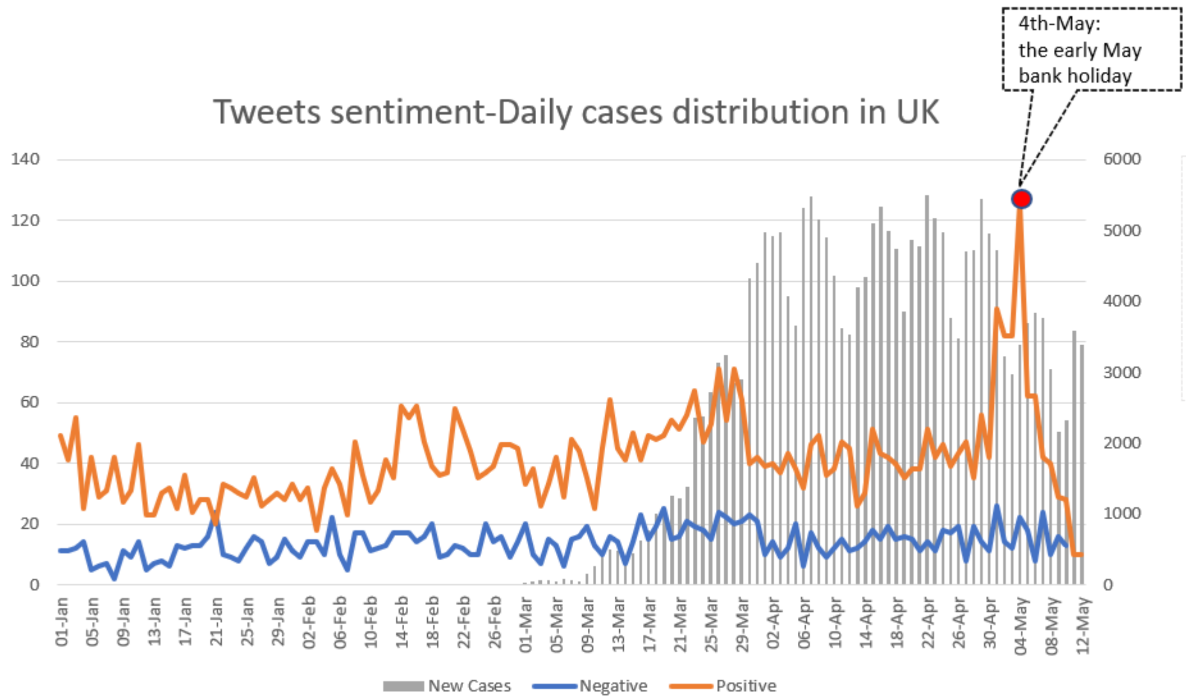
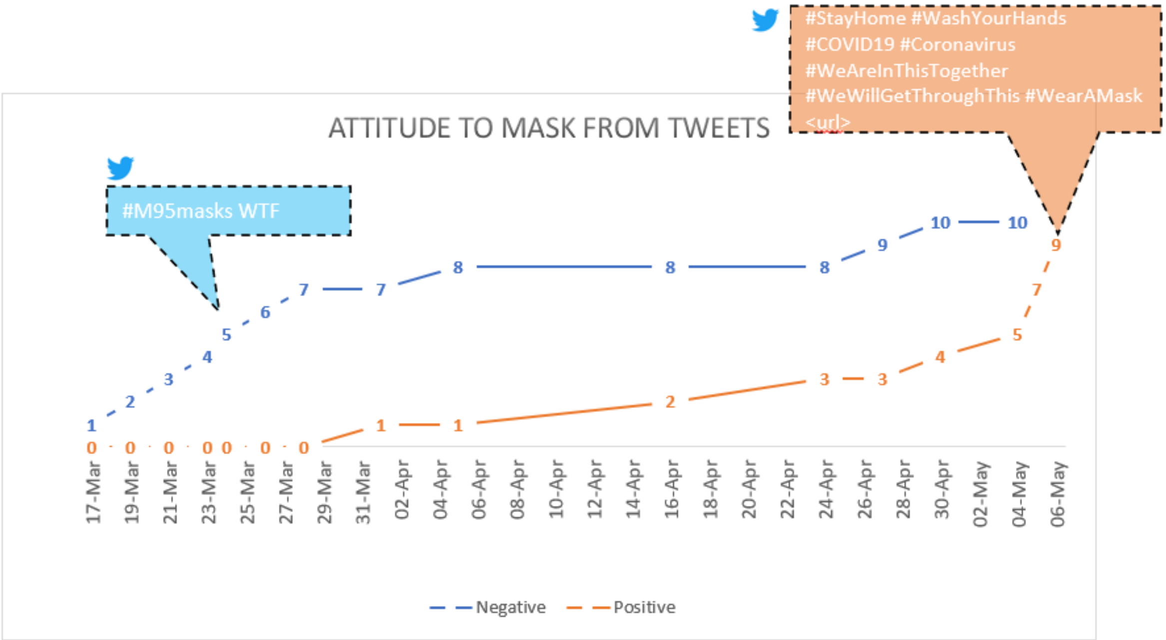


Figure 5.4 Tweets daily sentiment to keyword “mask”



Finally, we found that tweets which related to medical masks did not appear before March and they were negative in March but started positive from April, as shown in Figure 5.4.

6. CONCLUSION AND FUTURE WORKS

6.1 CONCLUSION

Equipped with tools we learned in this course from Professor Mageed and guided by our mentor Chiyu, we deployed, among others, a pre-trained BERTweet model to infer sentiments from a relatively small-sized tweets dataset, which consisted of 5 months ’data from 6 countries in 3 distinctive regions. We found that users’ general sentiments evolve with time, and the sentiments towards certain words, e.g., “(face) mask” change over time, finally there seem to be correlation between sentiments and outbreaks.

6.2 FUTURE WORKS

Admittedly, our project in its current form was limited by size of inference data and computing power at our disposal. If we were to continue our research, we could test our model with more data, both in terms of longer period and more geographic regions. We can also explore the data in greater graduality, e.g., comparing and contrasting the sentiments around more keywords, and continue to fine-tune our model’s hyper-parameters.

REFERENCE

[1] H. Manguri, K., N. Ramadhan, R. and R. Mohammed Amin, P. (2020) “Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks”, Kurdistan Journal of Applied Research, 5(3), pp. 54-65.

[2] Mathieu Cliché. “BB twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs”.

[3] C. Ziems, B. He, S. Soni, S. Kumar. “Racism is a Virus: Anti-Asian Hate and Counterhate in Social Media during the COVID-19 Crisis”.

[4] M. Abdul-Mageed, A. Elmadany, El M. Billah Nagoudi, D. Pabbi, K.Verma, R. Lin: Mega-COV: A Billion-Scale Dataset of 100+ Languages for COVID-19