SENWAVE: MONITORING THE GLOBAL SENTIMENTS UNDER THE COVID-19 PANDEMIC

Qiang Yang[a], Hind Alamro[a], Somayah Albaradei[a], Adil Salhi[a], Xiaoting Lv[b], Changsheng Ma[a], Manal Alshehri[a], Inji Jaber[c], Faroug Tifratene[a], Wei Wang[a,b], Takashi Gojobori[a], Carlos M. Duarte[d], Xin Gao[a] and Xiangliang Zhang✉[a]

[a]*Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division and Computational Biosciences Research Center (CBRC), King Abdullah University of Science and Technology (KAUST)*
Email: {qiang.yang, hind.alamro, somayah.albaradei, adil.salhi, changsheng.ma, manal.alshehri, faroug.tifratene, wei.wang, takashi.gojobori, xin.gao, xiangliang.zhang}@kaust.edu.sa
[b]*School of Computer Science and Information Technology, Beijing Jiaotong University, China*
Email: xiaotinglvtt@gmail.com
[c]*IT department, King Abdullah University of Science and Technology (KAUST)*
Email: inji.jaber@kaust.edu.sa
[d]*Red Sea Research Center (RSRC) and Computational Biosciences Research Center (CBRC), King Abdullah University of Science and Technology (KAUST)*
Email: carlos.duarte@kaust.edu.sa

June 22, 2020

ABSTRACT

Since the first alert launched by the World Health Organization (5 January, 2020), COVID-19 has been spreading out to over 180 countries and territories. As of June 18, 2020, in total, there are now over 8,400,000 cases and over 450,000 related deaths. This causes massive losses in the economy and jobs globally and confining about 58% of the global population. In this paper, we introduce SenWave, a novel sentimental analysis work using 105+ million collected tweets and Weibo messages to evaluate the global rise and falls of sentiments during the COVID-19 pandemic. To make a fine-grained analysis on the feeling when we face this global health crisis, we annotate 10K tweets in English and 10K tweets in Arabic in 10 categories, including *optimistic, thankful, empathetic, pessimistic, anxious, sad, annoyed, denial, official report*, and *joking*. We then utilize an integrated transformer framework, called *simpletransformer*, to conduct multi-label sentimental classification by fine-tuning the pre-trained language model on the labeled data. Meanwhile, in order for a more complete analysis, we also translate the annotated English tweets into different languages (Spanish, Italian, and French) to generated training data for building sentiment analysis models for these languages. SenWave thus reveals the sentiment of global conversation in six different languages on COVID-19 (covering English, Spanish, French, Italian, Arabic and Chinese), followed the spread of the epidemic. The conversation showed a remarkably similar pattern of rapid rise and slow decline over time across all nations, as well as on special topics like the herd immunity strategies, to which the global conversation reacts strongly negatively. Overall, SenWave shows that optimistic and positive sentiments increased over time, foretelling a desire to seek, together, a reset for an improved COVID-19 world.

Keywords Covid-19 · Pandemic · Sentimental analysis · Tweets · Weibo · Fine-grained sentiment annotation

1 Introduction

Since the outbreak of coronavirus, it has affected more than 180 countries where massive losses in the economy and jobs globally and confining about 58% of the global population are caused. Many people have been forced to work or study from home under pandemic. The research on people's feelings is essential for keeping mental health and informed about Covid-19. Social medias (e.g., Twitter, Weibo) have played a major role in expressing people's feelings

and attitudes towards Covid-19. We thus target on building a system named SenWave to monitor the global sentiments under the COVID-19 Pandemic by deep learning powered sentiment analysis.

Sentiment analysis has been widely researched in the field of natural language processing [1, 2, 3, 4]. Most of the current sentimental analysis tasks usually consider the coarse-grained emotion labels like *positive, neutral*, and *negative* for reviews/comments of books/products/movies, or five values to indicate the degree of emotions with the ranking score from 1 to 5. However, the feelings of people in pandemic are much more complicated than the sentiments in movie reviews and product comments ect. For instance, people may negatively feel *angry* and *sad* since Covid-19 leads to the increasing number of deaths and unemployment, while others may feel *optimistic* because of the medical supplies and medical assistance for the people in need. Therefore, we need to define the fine-grained labels to better understand the impact of the health crisis on sentiment.

However, there are no appropriate and sufficient annotated data to support the building of our SenWave by training deep learning sentiment classifiers. One non-Covid-19 tweet sentimental analysis dataset is avaliable in [5] with the labeled 7724 English tweets, 2863 Arabic tweets, and 4240 Spanish tweets (in total 14, 827), which is a benchmark dataset labeled in 11 categories, *anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise*, and *trust*. Initially, we tried to use it as our training data for building a Covid-19 sentiment classifier. However, the sentimental results were not suitable for Covid-19 analysis. For example, very few tweets are in the categories *joy, love* and *trust*. Besides, many tweets of *official* reports were classified into inappropriate categories as well as the tweets making *jokes* and *denying* conspiracy theories. Actually, these kinds of labels are essential in expressing opinions and attitudes according to our observations.

Due to the lack of appropriate training data, the Covid-19 tweet sentiment analysis has been conducted mostly based on engineered features or conventional bag-of-words derived representations not touching the deep learning models, either in unsupervised or supervised ways but with limited training data (e.g., only 5K in [6]). Yasin et al. built a real-time Covid-19 tweets analyzer using LDA model in USA data on *positive, neutral*, and *negative* [7]. Jia et al. used LDA model and NRC Lexicon on the English tweets to predict one of emotions from (*anger, anticipation, fear, surprise, sadness, joy, disgust* and *trust*) [8]. Mohammed et al. employed naïve Bayes model to predict Saudis' attitudes towards Covid-19 preventive measures on the (*postive, negative*, or *neutral*). Caleb et al. used logistic regression classifier with linguistic features, hashtags and tweet embedding to identify anti-Asian hate and counterhate text [9]. However, these methods are either limited by the sentimental dictionary and its availability or in lack of the deep understanding of tweets semantics. Even if the supervised methods are used like naïve Bayes, they still cannot satisfy the real case (multi-class classification) since emotions in Covid-19 can be a mixture of multiple emotions (multi-label classification).

In order to solve above problems, on one hand, we collected over 105+ million tweets covering six different languages, including English, Spanish, French, Arabic and Italian from March 1, 2020. On the other hand, to meet the requirement of fine-grained sentiment analysis target for Covid-19, we annotated 10,000 tweets in English and 10,000 tweets in Arabic in 10 categories, including *optimistic, thankful, empathetic, pessimistic, anxious, sad, annoyed, denial, official*, and *joking*. Each tweet was annotated by at least three experienced annotators in the corresponding language under strict quality control. We allowed one tweet to be annotated by more than one category, to support the multi-label analysis. To analyze over 1 million Chinese weibo of COVID-19 posts, we construct a set of 21,173 Weibo posts labeled in 7 sentimental categories, such as *anger, disgust, fear, optimism, sadness, surprised*, and *gratitude*. These 41K labeled datasets and the over 106+ million unlabeled tweets and weibo posts compose our dataset studied by SenWave. More details about dataset are shown in Sec. 2. We will make available the data and the SenWave implementation in public for supporting other social impact analysis of Covid-19 and fine-grained sentiment analysis. To comply with Twitter's Terms of Service, we will only publicly release the Tweet IDs for unlabeled data and limited number of tweet texts for labeled data (totally 20K) for non-commercial research use. To the best of our knowledge, this is the largest labeled Covid-19 sentimental analysis dataset with the fine-grained labels.

We summarize our main contributions in this paper as follows:

- We constructed the so far the largest fine-grained annotated Covid-19 tweets dataset (10K for English tweets and 10K for Arabic tweets) in 10 sentiment categories, which help to facilitate the studies of social impact of Covid-19 and other fine-grained analysis tasks in research community.

- We share a large set of Covid-19 tweets IDs collected since Mar 1, 2020, in five languages accumulated over 105+ million tweets, which will be continuously updated.

- We report the usability of the labeled Covid-19 tweets by first evaluating the performance of deep learning classifiers trained on them and then test them on the over 105 million unlabeled tweets from March 1 to May 15, 2020 to monitor how the global emotions vary in concerned topics and report other interesting findings.

This is the first report of COVID-19 sentiment over 105 million tweets. The largest analyzed COVID-19 tweet dataset before our work is 1.8M in [8], by an unsupervised way using topic modeling and lexicon features.

2 Dataset Construction

2.1 Data Collection

We used Twint[1], an open source Twitter crawler, to collect our tweet dataset where Twint allows users to specify a number of parameters alongside the query, such as tweet language, time period, etc. By forming requests with specified parameters, the resulting response was scraped into JSON documents. We used a unified query across these languages: "covid-19 OR coronavirus OR covid OR corona OR كورونا (corona in English) ". We launched 12 instances on 24 cores for downloading daily updates and historical data up to the March 1. Data rates varied slightly throughout the period averaging around a little over a million tweets a day. Tweets were saved as JSON documents, and pooled into a shared medium, to be pre-processed and consumed by the language models for sentiment analysis, running on a GPU server (GTX 1080ti GPU and 20 CPUs).

Due to the limited number of tweets in Chinese that could be found on Twitter, our data collection for the sentiment analysis of COVID-19 in China was conducted on Sina Weibo, which is the largest social media platform in China. The Weibo records were collected by Sina weibo API, starting from collecting first the hashtags about COVID-19, and then extracted weibo records including these hashtags.

2.2 Data Annotation

We randomly selected 10K English and 10K Arabic tweets for sentiment annotation. These two languages are selected because English and Arabic are among the top-5 popular languages in the world[2]. In addition, English can be effectively translated into other languages when needed. We thus have a considerably large labeled dataset for tweet sentimental analysis. The sentiment categories were determined by domain experts after reviewing a subset of the collected tweets and discussing for several rounds. The final determined set of labels reflect the complicated sentiments in pandemic. These 10 labels and their covered auxiliary emotions are *optimistic* (representing *hopeful, proud, trusting*), *thankful* for the efforts to combat the virus, *empathetic* (including *praying*), *pessimistic* (*hopeless*), *anxious* (*scared, fearful*), *sad*, *annoyed* (*angry*), *denial* towards conspiracy theories, *official* report, and *joking* (*ironical*).

We recruited over 50 experienced annotators to make every tweet labeled by at least three annotators. Example tweets were provided in advance to annotators with suggested categories. We allowed each tweet to be assigned to multiple labels, which is in line with the convention. For example, the tweet "Dear Covid19, Will you please retaliate on our behalf. We're hopeless, helpless, restless, speechless." has a mixture of *pessimistic* and *sad* emotions. We allowed the multi-label annotation with our fine-grained sentiments to support the analysis of the complicated emotions in the pandemic. In order to measure the reliability of the sentiment annotations, we conducted a verification study on the annotated tweets where the final results of the labeled tweets is determined by the majority-voting strategy. Our labeled dataset can also be used for other events with complex emotions, e.g., public opinion analysis and general election analysis. The commonality is that the emotions of these events are multiple aspects.

For Chinese weibo, we analyzed the COVID-19 posts and annotated 21,173 Weibo in 7 sentimental categories, such as *optimistic, thankful, surprised, fearful, sad, angry* and *disgusted*.

Table 1: Statics of data collected in each language

	En	Es	Ar	Fr	It	Zh
max	1.7M	652K	223K	132K	63K	23K
min	272K	90K	52K	28K	15K	0.6K
mean	898K	273K	105K	64K	35K	9.6K
total	68M	21M	8M	4.9M	2.7M	1.1M

Table 2: Correlations of normalized volume

	Es	Fr	It	Ar
En	0.94	0.96	0.80	0.83
Es		0.96	0.82	0.84
Fr			0.84	0.86
It				0.79

2.3 Data Description

The processed data is saved and updated in the git repository using the fetched data. To comply with Twitter's Terms of Service, we are only publicly releasing the Tweet IDs for unlabeled data and the limited number of Tweet texts for the label data. Hence, the researcher can refetch the original tweet if that tweet is still public. This dataset is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License (CC BY-NC-SA 4.0). The dataset is available on Github at `https://github.com/gitdevqiang/SenWave`.

---

[1] https://github.com/twintproject/twint

[2] https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets
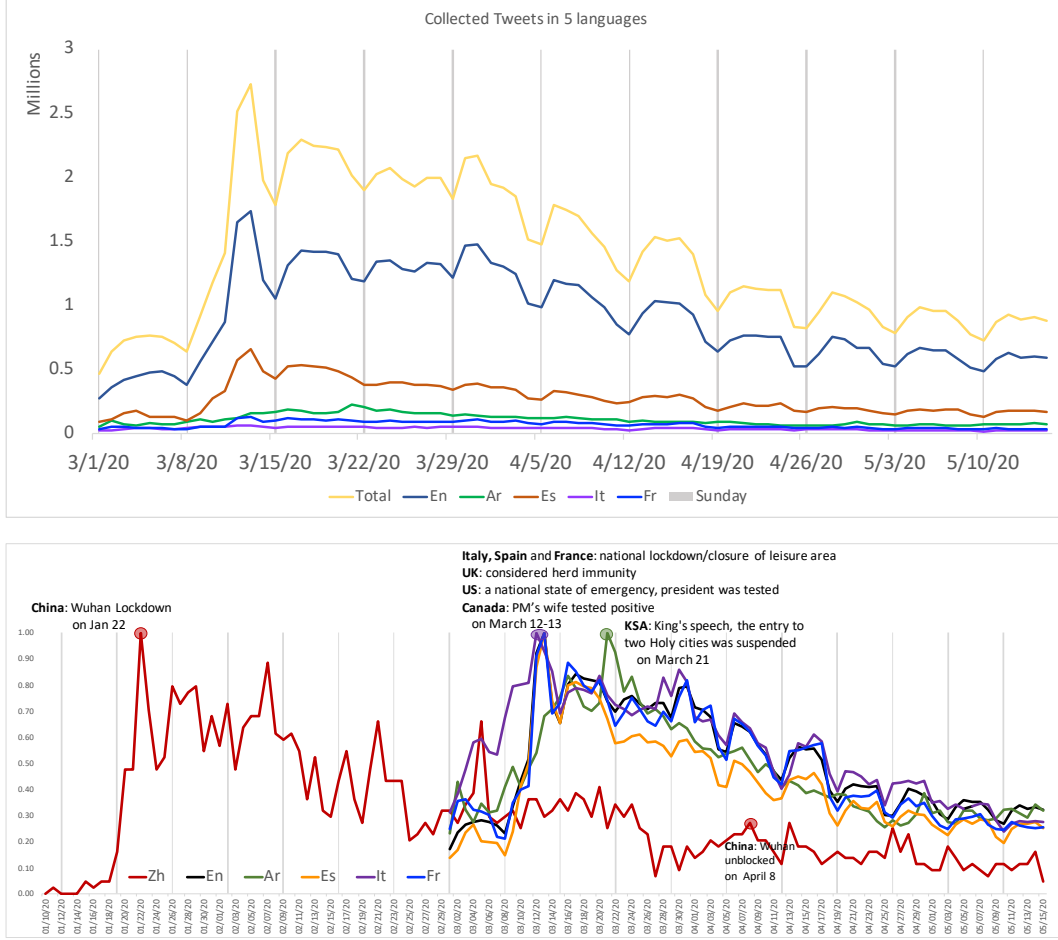
Figure 1: (Top) the absolute daiily volume of COVID-19 Tweets from Mar 1 to May 15, 2020 collected in 5 languages, English (En), Spanish (Es), Arabic (Ar), French (Fr), and Italian (It). (Bottom) the normalized volume (normalized to maximum number, with volume statistics in Table 1) of the 5 languages, as well as the COVID-19 Weibo in Chinese (Zh) from Jan 10 to May 15.

2.3.1 Statistics of Unlabeled Tweets

The volumes of collected daily data for each language are illustrated in Fig. 1. The statistics show a similar pattern of rapid rise followed by gradual fall in the global citizen conversation around COVID-19, but with the peak in messages in Chinese reaching the maximum on January 22, two months earlier than the peak in all other languages examined on March 12-13 and March 21, reflecting the lag between the development of the epidemic first detected in Wuhan and the spread to reach pandemic status. The discussion and attention quickly reached the peak when important decisions were made. For example, Italy, Spain, and France announced the national lockdown or closure of leisure areas while UK considered the herd immunity and US announced a state of alarm on March 12-13. Meanwhile, in KSA the peak was reached due to the King's speech and the suspended entry to two Holy cities on March 21. In addition, people's attention cools down as time goes on. As shown in Table 2, the fall and rise in the volume of messages on COVID-19 was remarkably correlated for English, French, Italian, Arabic and Spanish languages, although the rise occurred earlier in Italian, the first western nation to suffer the epidemic. The high correlation coefficient values indicate that populations speaking different languages are responding in a similar way.

2.3.2 Information of Annotated Tweets

The label distributions are shown in Table 3. Note that the sum of the percentages is not 1 due to the multi-label annotation in En and Ar. In English, *joking* and *annoyed* emotions took large portions, which is consistent with the the reality, since Covid-19 causes deaths, high unemployment rates and other problems. However, we also see *optimistic* emotion is the third largest category, indicating that people feel confident about combating the virus and about the

future. In Arabic, *official* is significantly higher than others, because since the outbreak of Covid-19, most of the Arabic governments announced a lot of decisions regarding different situations on Twitter.

Table 3: The label distributions in annotated English (En) and Arabic (Ar) datasets (%)

	Opti.	Than.	Empa.	Pess.	Anxi.	Sad	Anno.	Deni.	Offi.	Joki.
En	23.73	4.98	3.89	13.25	16.95	21.33	34.92	6.31	12.07	44.76
Ar	11.27	3.33	6.49	4.65	7.53	10.8	17.17	2.1	34.52	14.18

	Optimistic	Thankful	Surprised	Fearful	Sad	Angry	Disgusted
Zh	16.98	14.51	14.85	15.24	13.02	15.00	10.40

English tweet examples of each category are shown in Table 4 where some tweets have more than one label, even with three labels. Based on the statistics of categories, we find that more than 70% of English tweets were assigned with more than one label, while about 20% of Arabic tweets were assigned with more than one labels. We also present the relations between these labels in Fig. 2. While the label co-occurrence in Arabic shows as three blocks (positive, negative and neural), the label co-occurrence in English is more complicated. These observations imply that the multi-label classification in our English dataset is more challenging than that in Arabic. We also illustrate this fact in the experiments.
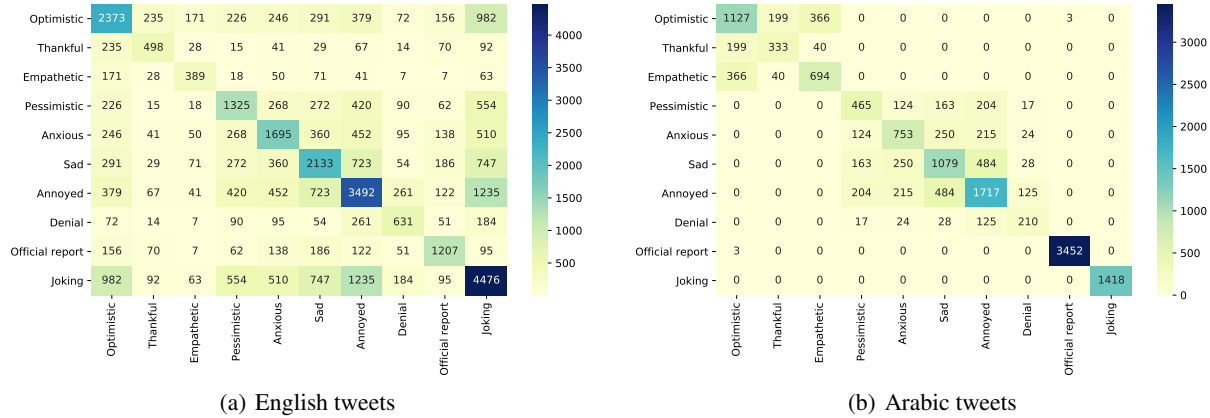

(a) English tweets   (b) Arabic tweets

Figure 2: Heatmaps of co-occurrence of labels for English and Arabic tweets.

Table 4: English examples of each category

Category	Example
	Single label
Optimistic	Nothing last forever, Corona Virus will Vanish this month. "Happy New Month"
Thankful	Gratitude to those who are involved to safeguard our lives from fatal Corona virus. Thanks to them. #LetUsPrayForCoronaFighters
Empathetic	Allah ap ko bhi safa ata fermain. Ameen. Be strong. IA Allah will give full and speedy recovery. #coronavirus
Pessimistic	things won't go back to normal #COVID19 #coronavirus #pandemic #MIT
Anxious	I don't feel good and I don't know if I'm just exhausted from working so much or if I have corona
Sad	When someone you know.. apart of your family dies from the Coronavirus it's shocking; unexplainable. My whole day has been down.
Annoyed	Stop asking to change location man hat how you will spread corona. Fooook
Denial	Unpopular and Insensitive Thought... Corona and Quarantine is a marketing campaign by OTT platforms...!!
Official report	Now schools in Ontario won't be open until May due to the Coronavirus which might post-pone the 2019-2020 school year to July or August.
Joking	Calling Corona Virus "rona" like she the nastiest little girl in the 5th grade. #coronavirusmemes #5G
	Multiple labels
Empathetic, Sad	So heart breaking any way you see it Prayers to all the families affected by the Covid-19.
Pessimistic, Joking	if i get curved ima go somewhere packed to give myself coronavirus
Anxious, Pessimistic	Does everyone realize we're going to reach a million cases of this coronavirus by the weekend?
Denial, Sad, Annoyed	Why is it that no one ever reports on the number of people who recovered from Coronavirus?
Joking, Annoyed	My uncle paranoid about corona virus but still goes to work .... pick one

3 Model Introduction

3.1 Data Preprocessing

We pre-processed the raw data to ensure the analysis quality. In details, we first remove the @users, and URLs from the tweet because they do not contribute to the tweet analysis. Then, we remove emojis and emoticons like ⌣ though they can express emotions well since we focused on the analysis textual data. Next, we filtered out noisy symbols and texts, which cannot convey meaningful semantic or lexicon information, and may even hinder the model from learning, such as retweet symbol "RT" and some special symbols including line break, tabs and redundant blank characters. Unlike previous methods which also removed hashtags in tweets, we kept these hashtags since they have meaningful semantics, like "Proud to be one of the few people who hasn't texted their ex #Covid-19 #Quarantine #lockdown". Apart from that, we also conducted word tokenization, steaming and tagging with the NLTK tool (https://www.nltk.org/) for English, Spanish, French and Italian, and with Pyarabic for Arabic (https://github.com/linuxscout/pyarabic). We used Jieba for Chinese weibo segmentation (https://github.com/fxsjy/jieba).

3.2 Multi-label Sentiment Classifiers

We built our multi-label sentiment classifiers based on deep neural network language models due to their success on diverse NLP tasks. An integration framework called simpletransformer (https://simpletransformers.ai/) supports the fine-tuning of these pre-trained models and the training of a customized classifier. We used XLNet [10] for English, AraBert [11] for Arabic, and ERNIE [12] for Chinese (selected due to the better performance of ERNIE than that of Bert [13] and LSTM).

We first evaluated the performance of the sentiment classifier in English and Arabic language on the 10K annotated tweets by 5-fold cross validation, except Chinese on 21K labeled Weibo. Then all 10K labeled tweets were used to train the final sentiment classifiers, except the Chinese sentiment classifier that was trained with 21K Chinese Weibo posts. The trained model was then used for predicting the sentiments of millions of Covid-19 tweets (Mar 1 - May 15, 2020 for Non-Chinese data and January 20 - May 15, 2020 for Chinese Weibo message) for our analysis.

Considering that the translation between English and Spanish, French, Italian has been well developed, we translated the labeled English tweets into Spanish, French, Italian with Google translate (https://translate.google.com/) to illustrate whether our classifiers can work well . We manually checked a subset of translated tweets and were surprised by the high quality of translation. We used Bert [13] for Spanish, French and Italian tweets representation learning, and then follow the same steps for sentiment analysis.

3.3 Experimental Setting and Evaluation Metrics

We ran the experiments on a workstation with one GeForce GTX 1080 Ti with memory size 11178MB. The batch size is 16, and the learning rate is $4e-5$ with 20 epochs. We used multi-label accuracy (Jaccard index), F1-macro, and F1-micro, as well as the weak accuracy as the performance metrics. The accuracy with Jaccard index is defined as:

$$Jac.Acc = \frac{1}{|D|} \sum_{i=1}^{D} \frac{Y_i \cap \hat{Y}_i}{Y_i \cup \hat{Y}_i}$$

where Y_i is the ground truth labels for the i-th testing sample, and $\hat{Y}_i$ is the predicted labels. And the weak accuracy of multi-label classification is defined as:

$$Acc = \frac{1}{D * m} \sum_{i=1}^{D} \sum_{j=1}^{m} \sigma(\hat{y}_{ij} == y_{ij})$$

where $\sigma(\hat{y}_{ij} == y_{ij})$ checks if the predicted $\hat{y}_{ij}$ is the same as the ground truth y_{ij}, which can be 1 meaning the i-th testing sample has a j-th label, and can be 0 indicating the i-th testing sample doesn't have j-th label. The total number of corrected prediction of y_{ij} is averaged with D by m, which are the number of testing samples and the number of labels, respectively. In addition, we used the ranking average precision score (LRAP) and Hamming loss, which are specified for multi-label classification.

4 Results and Analysis

4.1 Performance of Classifier

We present the 5-fold cross validation results of our multi-label classifiers on SenWave in Table 5. Our classifiers reach above 80% weak accuracy values, which prove the effectiveness of our models. The multi-label Jaccard accuracy of English and Arabic data is larger/equal than/to 0.5. However, the accuracy of Spanish, French and Italian tweets are not better than the original data. The reasons can be two-folds: 1) the usage of different pre-trained language models: XLNet used for English tweets and AraBert used for Arabic tweets perform better than Bert generally used for Spanish, French and Italian on the same conditions [10, 11]; 2) the difficulty of classifying multi-label English tweets due to the complex multi-labels (shown in Fig. 2 (a)). We are working on the improvement of these models. It is worth noting that F1 values are around 0.5 due to the class imbalance issue, which will be resolved in our future work. However, the high accuracy, LRAP and low Hamming loss demonstrate that the trained classifiers are usable for practical usage. The multi-class classification accuracy of Chinese weibo shown in Table 5 helped us select ERNIE (with accuracy 0.88) for the final analysis due to its better performance than Bert (with accuracy 0.83) and LSTM (with accuracy 0.78).

Table 5: Validation of the SenWave on the labeled records

	Acc	Jac.Acc	F1-Macro	F1-Micro	LRAP	Hamming Loss
English	0.847 ± 0.004	0.495 ± 0.008	0.517 ± 0.012	0.573 ± 0.008	0.745 ± 0.007	0.153 ± 0.004
Arabic	0.905 ± 0.002	0.589 ± 0.010	0.520 ± 0.018	0.630 ± 0.009	0.661 ± 0.009	0.111 ± 0.002
Spanish	0.823 ± 0.001	0.428 ± 0.004	0.434 ± 0.010	0.511 ± 0.003	0.684 ± 0.002	0.177 ± 0.001
French	0.824 ± 0.004	0.431 ± 0.011	0.431 ± 0.0116	0.510 ± 0.011	0.687 ± 0.006	0.176 ± 0.004
Italian	0.827 ± 0.002	0.438 ± 0.006	0.441 ± 0.011	0.516 ± 0.005	0.693 ± 0.005	0.173 ± 0.002
Chinese	0.88 ± 0.001 (ERNIE Acc)					

4.2 Sentiments Variation in Six Languages Over Days

We present the sentiments variation of 6 languages from March 1 to May 15, 2020 for Non-Chinese data and from January 20 to May 15, 2020 for Chinese Weibo message in Fig. 3. The statistics of these sentiment results are given in Table 6.

The sentiment results of English tweets are shown in Fig. 3 (a). All the positive emotions, including *optimistic, thankful* and *empathetic*, showed the similar trend of first rising up and then falling down. It implies that people first felt positive due to the various decisions made for combating the virus staring from the mid of March. However, the emotions went down in late April when a large number of people got infected. Among negative emotions, *anxious* and *joking* fell down with the slope -0.0004 and -0.0007 respectively as time went on, while the others went stable with slight changes. The *anxious* may be reduced by the increasing of medical supplies and the fact that people have known much better about Covid-19 than before and got used to the ways living with Covid-19. However, the resulted high unemployment rate and the high number of death may be the reason that *sad* and *annoyed* have been staying high.

The results of Arabic tweets shown in Fig. 3 (b) demonstrate significant variations in all categories of emotions. In particular, *optimistic* has been rising up, and *anxious, denial* and *joking* are falling down. The *sad* emotion keeps rising up due to the increasing number of new cases in several Arabic-speaking populations, such as Saudi Arabia, Qatar and United Arab Emirates (UAE).

The rise of *optimistic* and *thankful* and the fall of *pessimistic* and *annoyed* were also observed in Fig. 3 (c) of Spanish tweets. The similar trend of increase in *thankful* is observed in French tweets, as shown in Fig. 3 (d). However, the other emotions went stable, except the decline of *joking* and the sudden increase of *denial* to the conspiracy theory of lab source of corona-virus. Italian tweets also showed weak increase or decrease trends in most of the emotions, as shown in Fig. 3 (e), except those in *thankful* and *empathetic*.

The Chinese Weibo sentiments show strong variations, but no obvious trend of increase and decrease, as shown in Fig. 3 (f). The most significant decrease in *fearful* is observed in the very beginning on Jan 20, 2020, when human-to-human transmission was confirmed on that day. The fearful state continued until January 22, when Wuhan was locked, and the arrival of Chinese New Year. The significant jumping up of *sad* on April 4 due to the nationwide memorial for victims of Covid-19.

4.3 Sentiments Variation of Selected Areas Over Days

We selected some countries and areas to illustrate how the sentiments vary over days including USA, Washington D.C. , UK, Spain, Argentina and Saudi Arabia in Fig. 4. The statistics of these sentiment results are given in Table 7.

(a) English

(b) Arabic

(c) Spanish
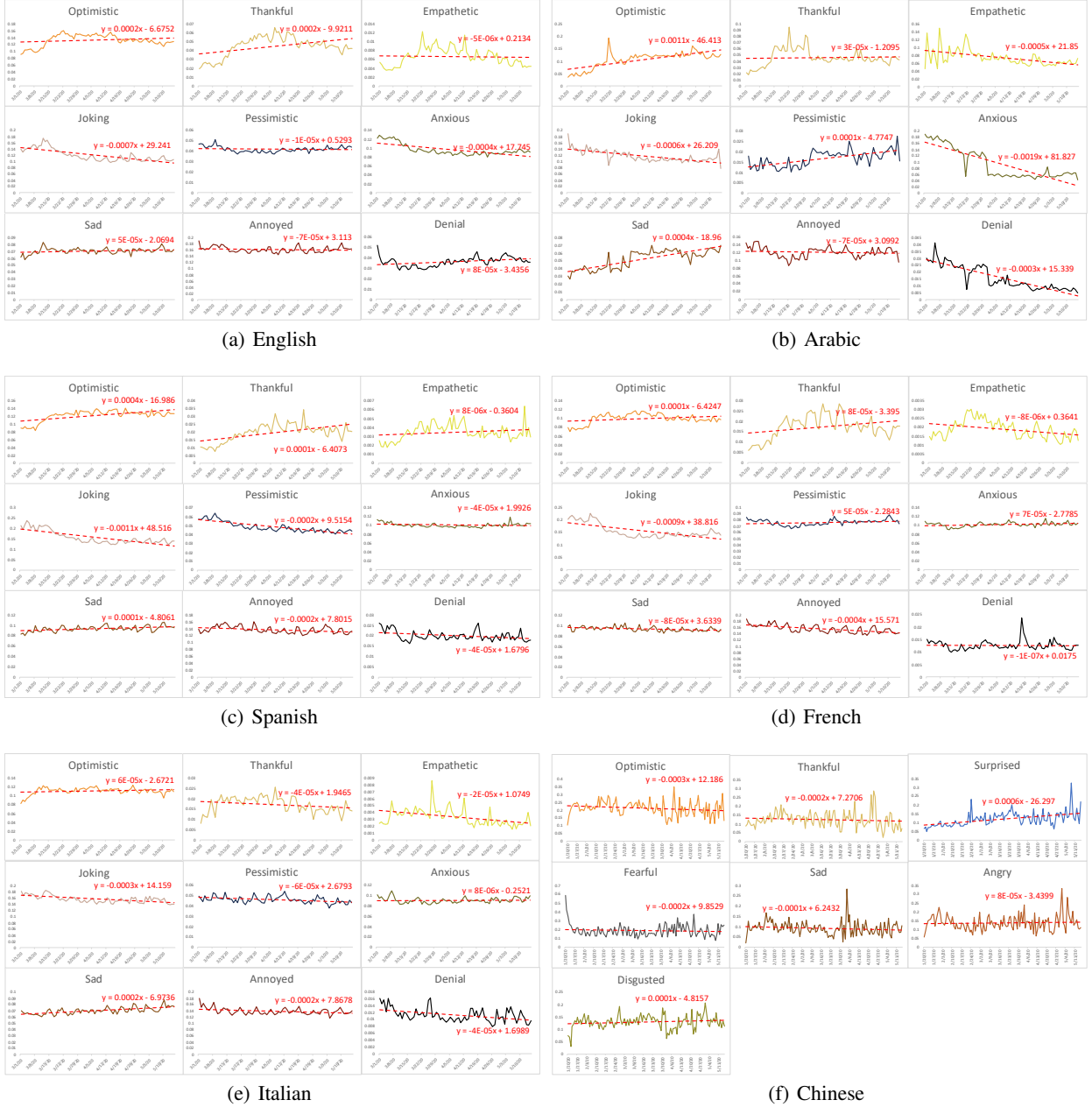
(d) French

(e) Italian

(f) Chinese

Figure 3: Sentiment variation of all languages over time. Each subfigure shows the results of one type of languages including 9 emotions for non-Chinese tweets and 7 emotions for Chinese Weibo messages. The linear regression line is fit to each emotion curve, showing the trend of the emotion variation.

Fig. 4 (a) shows in USA, the portion of negative emotions is higher than that of positive emotions. On March 12, people felt *annoyed* and *anxious* (see the pie charts) since the normal life was affected by coronavirus e.g., cancellation of sport events and suspending of transportation. On March 21, however, the positive emotions had a slight increase when people were showing gratitude for the efforts of healthcare workers. Similarly, on March 28, people were suggested to follow the guidelines to stay coronavirus free. However, the negative emotions went up once again due to the increasing rate of death, infection and unemployment on April 11. Several days later on April 18, this figures augmented because USA death toll was the highest in the world and it was still rising quickly.

Fig. 4 (b) shows in Washington D.C., on March 2, negative emotions (shown by the blue line) went to the peak due to a public health official announcement saying that outbreak of a new coronavirus could soon become a pandemic (see

Table 6: The statistics of daily sentiment fraction in different categories in all languages, presented as mean±std, and the number of tweets/weibo from which the statistics were obtained.

	Opti.	Than.	Empa.	Joki.	Pess.	Anxi.	Sad	Anno.	Deni.
En	0.19 ± 0.02 15758721	0.06 ± 0.02 5399197	0.01 ± 0.0 816664	0.17 ± 0.02 13822859	0.06 ± 0.0 4721744	0.14 ± 0.02 10872331	0.1 ± 0.01 8181085	0.23 ± 0.01 18445343	0.05 ± 0.01 4014464
Es	0.18 ± 0.02 4284423	0.03 ± 0.01 688944	0.0 ± 0.0 123807	0.22 ± 0.04 5472179	0.07 ± 0.01 1672356	0.14 ± 0.01 3410825	0.13 ± 0.01 3203353	0.2 ± 0.01 4759697	0.03 ± 0.0 671636
Ar	0.17 ± 0.05 844165	0.07 ± 0.02 392832	0.11 ± 0.03 627644	0.18 ± 0.02 943901	0.03 ± 0.01 127120	0.14 ± 0.06 794588	0.08 ± 0.02 399315	0.19 ± 0.02 941615	0.02 ± 0.01 138903
Fr	0.14 ± 0.02 884553	0.02 ± 0.01 158938	0.0 ± 0.0 17301	0.22 ± 0.03 1351669	0.11 ± 0.01 650069	0.14 ± 0.01 874491	0.13 ± 0.01 814349	0.22 ± 0.01 1357922	0.02 ± 0.0 108614
It	0.17 ± 0.01 486093	0.03 ± 0.01 78086	0.01 ± 0.0 15332	0.24 ± 0.01 683951	0.07 ± 0.0 201141	0.14 ± 0.01 391758	0.11 ± 0.01 300875	0.22 ± 0.01 604932	0.02 ± 0.0 49141

	Optimistic	Thankful	Surprised	Fearful	Sad	Angry	Disgusted
Zh	0.21 ± 0.05 244472	0.12 ± 0.04 143374	0.12 ± 0.04 124820	0.19 ± 0.07 206767	0.09 ± 0.03 105974	0.14 ± 0.04 155083	0.13 ± 0.03 143833

Table 7: The statistics of daily sentiment fraction in different categories of different countries and areas, presented as mean±std, and the number of tweets from which the statistics were obtained.

	Opti.	Than.	Empa.	Pess.	Anxi.	Sad	Anno.	Deni.	Offi.	Joki.
USA	0.16±0.02 271961	0.05±0.01 83907	0.01 ± 0.0 15011	0.04 ± 0.0 67987	0.11±0.02 179757	0.09±0.01 157241	0.2 ± 0.01 318295	0.03±0.01 48253	0.12±0.01 184803	0.18±0.02 308588
D.C.	0.17±0.03 2178	0.07±0.02 816	0.01±0.01 84	0.04±0.01 551	0.1 ± 0.03 1305	0.08±0.02 1064	0.18±0.02 2349	0.03±0.01 377	0.17±0.03 2195	0.14±0.02 1846
UK	0.2 ± 0.03 34834	0.07±0.02 13220	0.01±0.01 1574	0.05±0.01 7525	0.12±0.02 19233	0.1 ± 0.01 16606	0.17±0.02 27761	0.03±0.01 3977	0.11±0.02 17611	0.15±0.03 25245
Spain	0.18±0.03 4354	0.03±0.01 669	0.01 ± 0.0 145	0.06±0.01 1556	0.11±0.01 2807	0.13±0.02 3255	0.18±0.02 4560	0.02±0.01 567	0.09±0.02 2246	0.19±0.04 4620
Arge.	0.15±0.03 2938	0.02±0.01 304	0.0 ± 0.0 56	0.07±0.02 1452	0.12±0.02 2381	0.11±0.02 2277	0.18±0.02 3654	0.03±0.01 504	0.09±0.03 1708	0.23±0.04 4897
KSA	0.19±0.06 15137	0.09±0.03 7943	0.14±0.05 18162	0.02±0.01 1269	0.1 ± 0.06 8171	0.06±0.01 4623	0.12±0.03 6597	0.02±0.01 385	0.16±0.06 13261	0.11±0.03 6804

the pie chart at the right hand). On March 13, there was a positive spike due to the declaration of a National Day of Prayer amid coronavirus pandemic. On March 29, negative pumped up due to the high number of death within one day achieving 2000, and especially a significant increase of *anxious* because of the negligence of social distancing, even having gathering. On April 13, many people showed gratitude to the medical professionals and front-line workers (see the pie chart at the right hand) while during April 27 and 28, people expressed their dissatisfaction with the government due to the resistance, as well as on May 10.

Fig. 4(c) shows in UK, on March 9, the negative emotions caused by panic buying of hand sanitizer and toilet roll and people's fear to coronavirus and oil price war leading to the plunging of FTSE 100. After different cornoavirus measures were imposed, the positive sentiment went up significantly. It would be better to zoom in the figures to see other detailed interpretation.

In Spain (Fig. 4(d)), people applauded for the health care workers treating the coronavirus on the balcony on March 15, felt angry about the extension of another 15 days of alarm and sad about the third highest number of deaths on March 22 (in pie chart).

In Argentina shown in Fig. 4 (e), the proportion of negative emotions was very close to 0.5 even much higher in some days. On March 8, the discussions about first dead case of coronavirus and dengue were focused on leading to the increase of *anxious*, *sad* and *annoyed* (see pie chart at the right hand). On March 21, the feelings of stress, anxiety, panic went up because of the long quarantine, which resulted in the increase of *anxious* and *sad*. On April 29, more than 2,300 prisoners were released because of the coronavirus, which increased the feelings of *pessimistic*, *anxious* and *annoyed*.

Fig. 4 (f) shows stronger positive sentiment in Saudi Arabia than in other countries or areas. Especially, starting from March 13, there was an increase of positive emotions when a lot of decisions were taken by the Saudi government. The peak was reached on March 21, responding to a tweet by the Saudi minister of health: "We are all responsible, staying home is our strongest weapon against the virus". Another positive peak was shown on April 23-24, when Ramadan started.

(a) USA

(b) Washington D.C.

(c) UK

(d) Spain

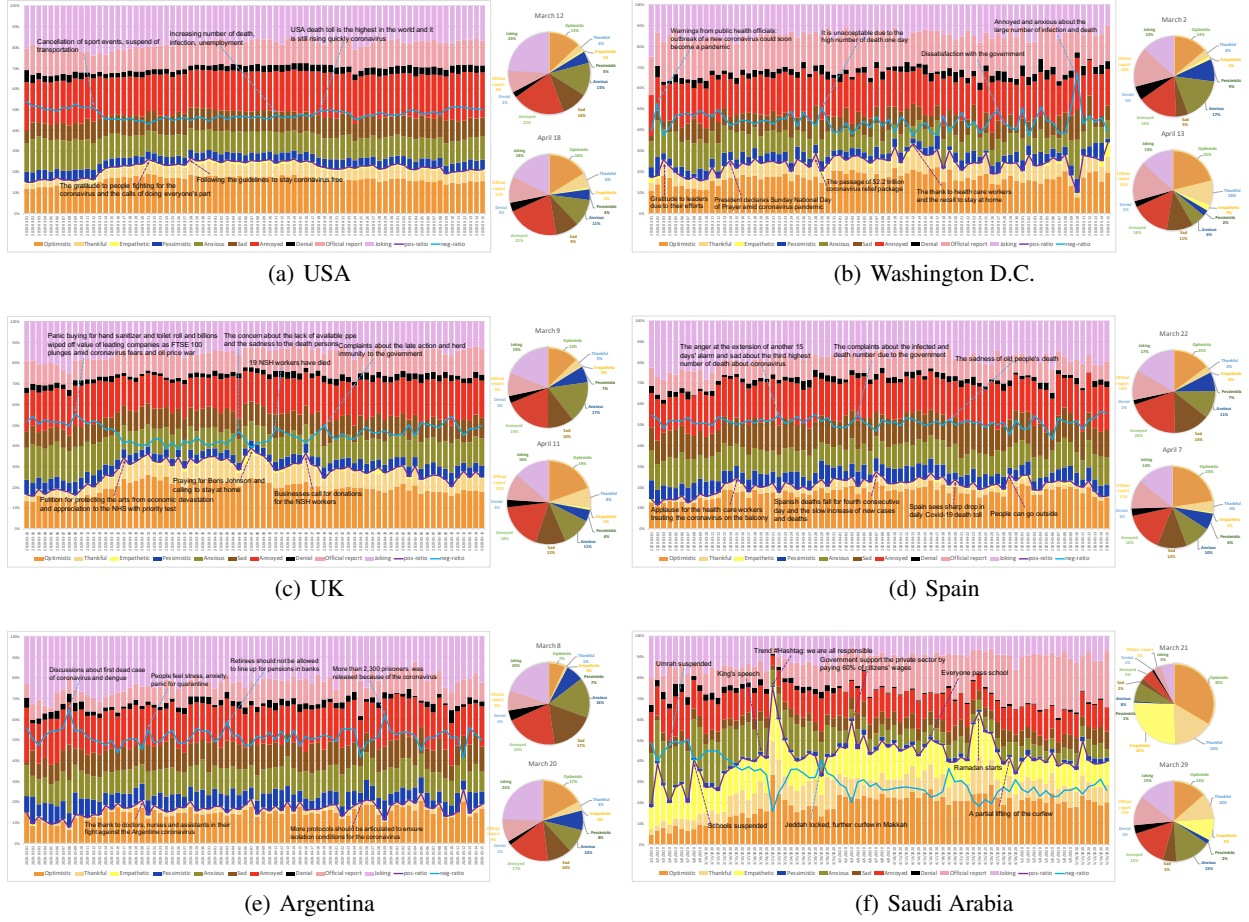(e) Argentina

(f) Saudi Arabia

Figure 4: Sentiment variation in different regions over time. Each bar shows the distribution of sentiments on one day, where sentiments are shown in different colors. The blue curve and purple curve show the positive (sum of *optimistic, thankful, empathetic* in yellow at different intensity) and the negative (sum of *pessimistic, anxious, sad, annoyed, denial* in blue at different intensity), respectively. (Better zoom in to see the interpretation of spikes)

4.4 Sentiments Variation of Studied Topics Over Days

We give the sentiment analysis of 7 topics including *stock market*, *oil price*, *herd immunity*, *economic stimulus*, *drug/medicine/vaccine*, *employment/job*, and *working from home*. The results are shown in Figure 5. The statistics of these sentiment results are given in Table 8.

The topic stock markets collapsed on March 9, when the peak of discussion was reached. On this day, *anxious* reaches a high value, which is greater than mean+2*std (out of the black dash line, and the black line is the mean, the dot line is the mean-2*std). On March 12, the DJI (Dow Jones Index) had its worst day since 1987, plunging about 10% (the second time breakers) and the volumes arrived the second largest. The *anxious* state remained at a high rate during these days. On the weekends of March 20-21 and March 28-29, the spikes of *denial* are higher than the blue dash line (mean+2*std), as a reflection of the continuous stock market collapse. DJI: Dow Jones Industrial Average is shown on top of the *denial* curve, and the markers indicate the days of March 9, 20 and 28, when these negative sentiments are reflected on the drop of DJI, as shown in Fig.5(a).

The topic oil price also showed the peak of discussion on March 9. The drop of crude oil price resulted in significant *anxious* on March 9-12. However, this was not the worst. On April 21, crude oil price reached an 18-year low, which is shown on the marked point on the WTI crude oil curve. Among the triggered discussion, we see *pessimistic* was significant.

The topic herd immunity quickly reached the top on March 14-15 when UK government initially considered it on March 13. Among the intensive discussions from March 13 to 17, *denial* and *joking* were significantly observed on

10

(a) Stock market

(b) Oil price

(c) Herd immunity

(d) Economic stimulus

(e) Drug/medicine and vaccine
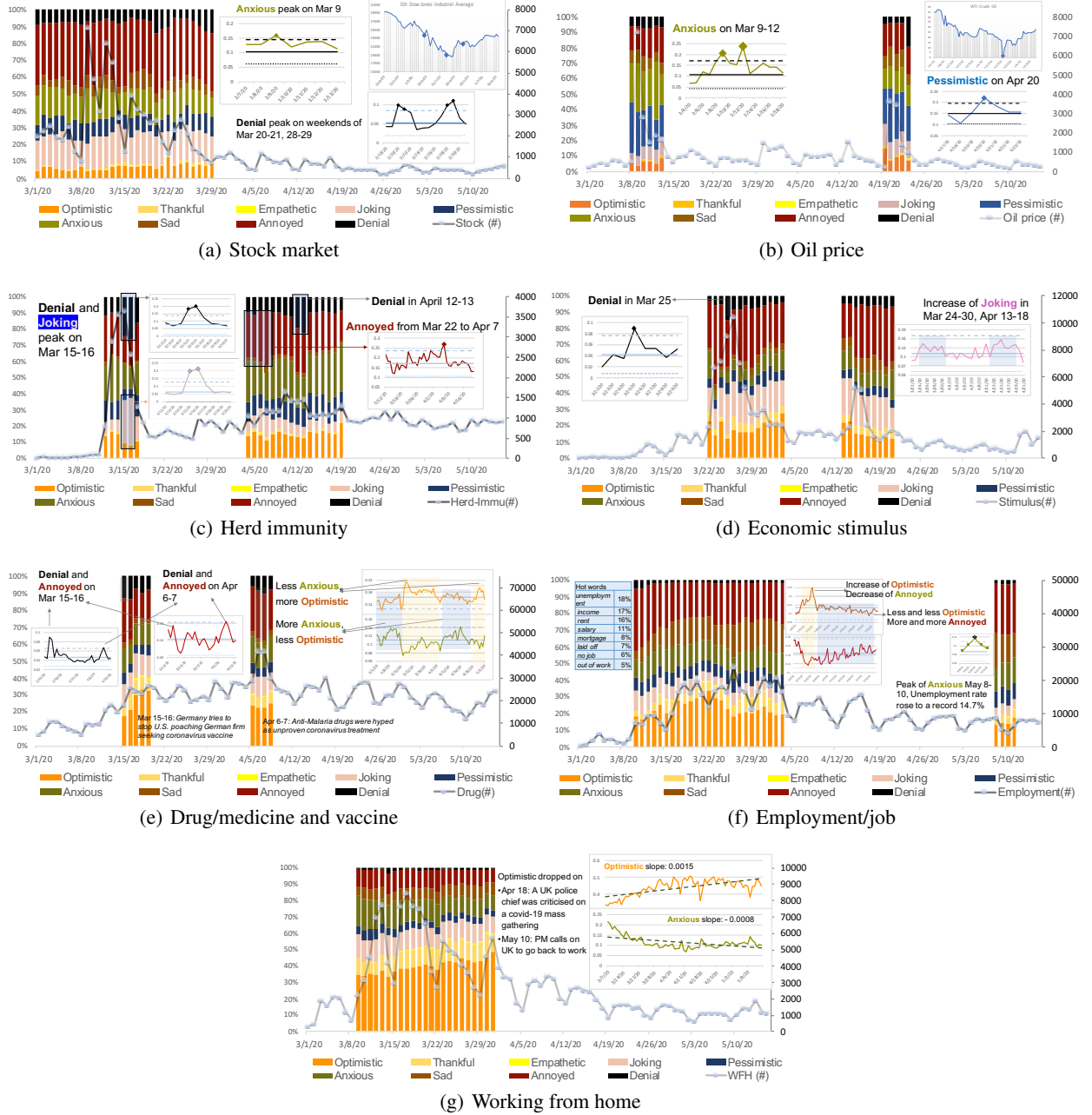

(f) Employment/job

(g) Working from home

Figure 5: Sentiments variation on seven topics. We show the sentiment results for these topics when they were intensively discussed (around the peak of volume curve in the background).

March 15-16. The discussion continued with significant *annoyed* from March 22 to April 7, and caused another rise of *denial* on April 12-13.

The topic economic stimulus reached top on March 26 when US Senate passed historic $2tn relief package. And another peak on April 15-16 when the checks were received. Surprisingly, during the discussion in March 23-26, positive was lower compared to other days, and *denial* was significant on March 25. We found many tweets under this topic are for example "this is not enough", "US economy is tanking", and "the pandemic is getting worse". By looking into the *joking*, we see increases in March 24-30 and April 13-18.

The topic drug/medicine/vaccine collected the largest amount of discussion among these 7 topics (reaching 20-40K on the daily volume). This topic has been hot since the global outbreak around March 10. Two events in this topic caused

11

significant *denial* and *annoyed*. The first event was on March 15-16, Germany tries to stop U.S. poaching German firm seeking coronavirus vaccine. The second event was on April 6-7, when Anti-Malaria drugs were hyped as unproven coronavirus treatment. Overall from March to May, we see two sections of more *anxious* and less *optimistic*, and two other sections of less *anxious* and *optimistic*.

The topic employment/job covered the hot words such as unemployment, income, rent, salary, mortgage, laid off, no job/work etc, as shown in the table included in Fig.5(f). In March, we see the increase of *optimistic* and the decrease of *annoyed*, however in April-May, we see less *optimistic* and the increase of *annoyed*. Peak of *anxious* was found on May 8-10, when the reported April unemployment rate rose to a record 14.7% in US.

The topic working from home (WFH) is a warm topic, which obviously has more positive than all other topics. The bars in Fig.5(g) show that optimistic took more than 40% in WFH, which is much higher than optimistic in other topics (between 10% and 25%) shown in Fig.5(a-f). We also see that *optimistic* keeps increasing with a slope of 0.0015, while *anxious* is decreasing with a slope of -0.0008. Only on two days, *optimistic* dropped: April 18 when a UK police chief was criticised on a covid-19 mass gathering, and May 10 when PM calls on UK to go back to work.

Table 8: The statistics of daily sentiment fraction in different categories under different topics, presented as mean±std, and the number of tweets from which the statistics were obtained.

	Opti.	Than.	Empa. (10^{-2})	Joki.	Pess.	Anxi.	Sad	Anno.	Deni.
Stock	0.09 ± 0.03 10875	0.01 ± 0.01 31970	0.08 ± 0.09 138	0.11 ± 0.02 26890	0.14 ± 0.03 15147	0.07 ± 0.02 22376	0.32 ± 0.05 10325	0.07 ± 0.02 48110	0.19 ± 0.03 10333
Oil	0.10 ± 0.03 6862	0.01 ± 0.01 909	0.18 ± 0.17 141	0.25 ± 0.04 7325	0.18 ± 0.05 20122	0.1 ± 0.03 15330	0.2 ± 0.05 7546	0.06 ± 0.02 14755	0.1 ± 0.02 4352
Herd Imm.	0.18 ± 0.05 17698	0.01 ± 0.01 1250	0.19 ± 0.55 108	0.13 ± 0.03 11605	0.24 ± 0.05 13723	0.02 ± 0.01 26275	0.23 ± 0.04 2335	0.1 ± 0.04 27006	0.09 ± 0.04 13865
Econ. Stim.	0.19 ± 0.05 31079	0.05 ± 0.02 8717	0.13 ± 0.25 169	0.07 ± 0.03 33923	0.09 ± 0.03 11270	0.05 ± 0.03 14986	0.3 ± 0.05 8586	0.07 ± 0.03 55648	0.18 ± 0.04 12483
Drug	0.25 ± 0.03 537330	0.07 ± 0.01 154398	0.50 ± 0.15 10749	0.07 ± 0.01 249193	0.17 ± 0.02 143002	0.03 ± 0.0 347756	0.22 ± 0.03 70442	0.07 ± 0.01 461216	0.12 ± 0.02 150892
Job	0.2 ± 0.05 209210	0.06 ± 0.02 61106	0.41 ± 0.18 4489	0.09 ± 0.01 68115	0.13 ± 0.01 84671	0.15 ± 0.04 127257	0.28 ± 0.04 156168	0.02 ± 0.01 255922	0.07 ± 0.02 20443
WFH	0.43 ± 0.05 125176	0.1 ± 0.02 31970	0.28 ± 0.17 975	0.05 ± 0.01 32971	0.12 ± 0.04 14188	0.07 ± 0.01 36042	0.1 ± 0.02 21320	0.01 ± 0.01 30036	0.11 ± 0.02 3271

5 Conclusion

This paper contributed a Covid-19 sentimental analysis system with annotate datasets, called SenWave, which includes 20K labeled English and Arabic tweets, and 21K labeled Chinese Weibo, as well as 106M+ Covid-19 tweets and Weibo messages collected since Mar 1, 2020 and January 20 respectively. We trained classifiers for 6 languages based on deep learning language models to monitor the global sentiments under the Covid-19 pandemic sentimental. We analyzed the sentiments varying of all languages and hot topics over days. On one hand, the emotions on these languages can be directly reflected by corresponding events at the specific days through the varying of volumes and a significant increase of emotions. On the other hand, the sentimental varying trends of 7 topics are also analyzed by showing the corresponding emotions. This work helps to provide a rich resource to the community to study and combat COVID-19.

References

[1] Ansari Fatima Anees, Arsalaan Shaikh, Arbaz Shaikh, and Sufiyan Shaikh. Survey paper on sentiment analysis: Techniques and challenges. Technical report, EasyChair, 2020.

[2] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253, 2018.

[3] Vishal Kharde, Prof Sonawane, et al. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*, 2016.

[4] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

[5] Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, 2018.

[6] Bennett Kleinberg, Isabelle van der Vegt, and Maximilian Mozes. Measuring emotions in the covid-19 real world worry dataset. *arXiv preprint arXiv:2004.04225*, 2020.

[7] Md Kabir, Sanjay Madria, et al. Coronavis: A real-time covid-19 tweets analyzer. *arXiv preprint arXiv:2004.13932*, 2020.

[8] Jia Xue, Junxiang Chen, Chen Chen, ChengDa Zheng, and Tingshao Zhu. Machine learning on big data from twitter to understand public reactions to covid-19. *arXiv preprint arXiv:2005.08817*, 2020.

[9] Caleb Ziems, Bing He, Sandeep Soni, and Srijan Kumar. Racism is a virus: Anti-asian hate and counterhate in social media during the covid-19 crisis. *arXiv preprint arXiv:2005.12423*, 2020.

[10] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5754–5764, 2019.

[11] Wissam Antoun, Fady Baly, and Hazem Hajj. Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*, 2020.

[12] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*, 2019.

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

Appendix

We present the hot words of the predicted English tweets for each category shown in Fig. 6 on March 9, 2020. More representations of different languages will be provided in future. The class *optimistic* is represented with hand washing and health, which means people should wash their hands frequently to keep health. The class *thankful* is presented with Covid-19 testing, while the class *empathetic* is shown with pray, hope, god and safe. The class *pessimistic* is reflected with economy market, oil market and large number of death. These hot words are also suitable for the class *anxious*. People felt *sad* about a lot of death and confirmed cases and the lockdown of school. The class *annoyed* is displayed with dont and flu while the class *denial* is demonstrated with market and China since some people didn't believe the Covid-19 report of China. Overall, these hot words in each category can represent the sentiments to some extend.

(a) Optimistic   (b) Thankful   (c) Empathetic   (d) Pessimistic

(e) Anxious   (f) Sad   (g) Annoyed   (h) Denial

(i) Official report   (j) Joking

Figure 6: Hot words of each category for English tweets on March 9, 2020