



Article

Large-Scale, Language-Agnostic Discourse Classification of Tweets During COVID-19

Oguzhan Gencoglu

Faculty of Medicine and Health Technology, Tampere University, 33720 Tampere, Finland;
oguzhan.gencoglu@tuni.fi

Received: 31 October 2020; Accepted: 27 November 2020; Published: 29 November 2020



Abstract: Quantifying the characteristics of public attention is an essential prerequisite for appropriate crisis management during severe events such as pandemics. For this purpose, we propose language-agnostic tweet representations to perform large-scale Twitter discourse classification with machine learning. Our analysis on more than 26 million coronavirus disease 2019 (COVID-19) tweets shows that large-scale surveillance of public discourse is feasible with computationally lightweight classifiers by out-of-the-box utilization of these representations.

Keywords: text classification; sentence embeddings; Twitter; natural language processing; deep learning; health informatics; COVID-19

1. Introduction

Coronavirus disease 2019 (COVID-19) was declared a pandemic by the World Health Organization on 11 March 2020 [1]. Since the first recorded case in Wuhan, China in late December 2019, 45.6 million people have been infected by COVID-19 and consequently, 1.2 million people have lost their lives globally as of 30 October 2020 [2]. This constitutes 700 times more deaths than SARS and MERS combined [3]. During such large-scale adverse events, monitoring information seeking behavior of citizens, understanding general overall concerns, and identifying recurring discussion themes is crucial for risk communication and public policy making [4,5]. This need is further amplified in a global pandemic such as COVID-19 as the primary responsibility of risk management is not centralized to a single institution, but distributed across society. For instance, a recent study by Zhong et al. shows that people's adherence to COVID-19 control measures is affected by their knowledge and attitudes towards the disease [6]. Previous national and global adverse health events show that social media surveillance can be utilized successfully for systematic monitoring of public discussion due to its instantaneous global coverage [7–12].

Twitter, due to its large user-base, has been the primary social media platform for seeking, acquiring, and sharing information during global adverse events, including the COVID-19 pandemic [13]. Especially during the early stages of the global spread, millions of posts have tweeted in the span of a couple of weeks [14–18]. Consequently, several studies proposed and utilized Twitter as a data source for extracting insights on public health as well as insights on public attention during the COVID-19 pandemic. Focus of these studies include nowcasting or forecasting of the disease, sentiment analysis, topic modeling, and quantifying misinformation/disinformation. Due to the novelty and unknown epidemiological characteristics of COVID-19, accurate quantification of public discussions on social media becomes especially relevant for disaster management (e.g., devising timely interventions or clarifying common misconceptions).

So far, manual or automatic topical analyses of discussions on Twitter during the COVID-19 pandemic have been performed in an exploratory and descriptive manner [19–21]. Characterizing public discourse in these studies relies predominantly on manual inspection, aggregate statistics of

keyword counts, or unsupervised topic modeling by utilizing joint distributions of word co-occurrences followed by qualitative assessment of discovered topics. One reason for previous studies to avoid supervised machine learning approaches might be the lack of annotated (labeled) datasets of public discourse on COVID-19. Furthermore, previous studies either restrict their scopes to a single language (typically English tweets) or examine tweets from different languages in separate analyses. This is mainly due to limitations of traditional topic modeling algorithms as they usually do not operate in a multilingual or cross-lingual fashion.

In this study, we propose large-scale characterization of public discourse themes by categorizing more than 26 million tweets in a supervised manner, i.e., classifying text into semantic categories with machine learning. For this purpose, we utilize two different annotated datasets of COVID-19 related questions and comments for training our algorithms. To be able to capture themes from 109 languages in a single model, we employ state-of-the-art multilingual sentence embeddings for representing the tweets, i.e., Language-agnostic BERT Sentence Embeddings (LaBSE) [22]. Our results show that large-scale surveillance of COVID-19 related public discourse themes and topics is feasible with computationally lightweight classifiers by out-of-the-box utilization of these representations. We released the full source code of our study along with the instructions to access the experiment datasets (https://github.com/ogencoglu/Language-agnostic_BERT_COVID19_Twitter). We believe our work contributes to the pursuit of expanding social media research for disaster informatics regarding health response activities.

2. Relevant Work

2.1. COVID-19 Twitter

Content analysis of Twitter data has been performed by various studies during the COVID-19 pandemic. Some studies approach their research problem by manual or descriptive (e.g., n-gram statistics) content analysis of Twitter chatter for gaining relevant insights [21,23–29], while other studies utilize unsupervised computational approaches such as topic modeling [19,20,30–49]. A high percentage of studies performing topic modeling and topic discovery on Twitter utilize the well-established Latent Dirichlet Allocation (LDA) algorithm [20,30,33,34,36,37,40–46,49]. Similar unsupervised approaches of word/n-gram clustering [38,39,47] or clustering of character/word embeddings [35,48] have been proposed as well.

Tweet data utilized for most of these studies are restricted to a single language. Majority of the studies restrict their analysis only to English tweets [19,24,29,30,33,37,39,41,43–46,48], possibly exacerbating the already existing selection bias. Other studies have restricted their datasets to Japanese [47], Korean [21], Persian/Farsi [36], and Polish [31] tweets. While studies that collect multilingual tweets exist, they have conducted their analyses (e.g., topic modeling) separately for each language [23,25,40].

2.2. Representing Tweets

As effective representation learning of generic textual data has been studied extensively in natural language processing research, tasks involving social media text benefit from recent advancements as well. While traditional feature extraction methods relying on word occurrence counts (e.g., *bag-of-words* or *term frequency-inverse document frequency*) have been extensively utilized in previous studies involving Twitter [50–52], they have been replaced by distributed representations of words in a vector space (e.g., *word2vec* [53] or *GloVe* [54] embeddings). Distributed word representations are learned from large corpora by a neural network, resulting in words with similar meanings being mapped to closer vector representations with a feature number that is much smaller than the vocabulary size. Consequently, sentences, documents, or tweets can be represented, e.g., as an average-pooling of its word embeddings. Such representations have also been learned specifically from Twitter corpora as *tweet2vec* [55,56] or *hashtag2vec* [57].

While distributed word/sentence embeddings provide effective capturing of semantics, they operate as a static mapping from the textual space to the latent space. Serving essentially as a dictionary look-up, they often fail to capture the context of the textual inputs (e.g., polysemy). This drawback has been circumvented by contextual word/token embeddings such as *ELMo* [58] or *BERT* [59]. Contextual word embeddings enable the possibility of the same word being represented as different vectors if it appears in different contexts. Several studies involving tweets utilized these deep neural network techniques or their variants either as a pre-training for further downstream tasks (e.g., classification, clustering, entity recognition) or for learning tweet representations from scratch [60–67]. Even though BERT word embeddings are powerful as pre-trained language models for task-specific fine-tuning, Reimers et al. show that out-of-the-box sentence embeddings of BERT and its variants can not capture semantic similarities between sentences, requiring further training for that purpose [68]. They propose a mechanism for learning contextual sentence embeddings using BERT neural architecture, i.e., *sentence-BERT*, enabling large-scale semantic similarity comparison, clustering, and information retrieval with out-of-the-box vector representations [68]. Studies involving Twitter data have been utilizing these contextual sentence embeddings successfully as well [69–72].

3. Methods

3.1. Data

For Twitter data, we utilized the publicly available dataset of 152,920,832 tweets (including retweets) related to COVID-19 between the dates 4 January 2020–5 April 2020 [73]. Tweets have been collected using the Twitter streaming API with the following keywords: *COVID19*, *CoronavirusPandemic*, *COVID-19*, *2019nCoV*, *CoronaOutbreak*, *coronavirus*, *WuhanVirus*, *covid19*, *coronaviruspandemic*, *covid-19*, *2019ncov*, *coronaoutbreak*, *wuhanvirus* [74]. These keywords capture the bulk of the Twitter chatter related to COVID-19 for the analysis time-frame according to [74]. As Twitter Terms of Service does not allow redistribution of tweet contents, only tweet IDs are publicly available. Extraction of textual content of tweets, timestamps, and other meta-data was performed with the use of open-source software Hydrator (<https://github.com/DocNow/hydrator>) with a Twitter developer account. For our study, we discard the retweets and at the time of extraction 26,759,164 unique tweets were available which is the final number of observations used in this study. Daily distribution of these tweets (7-day rolling average) can be observed from Figure 1.

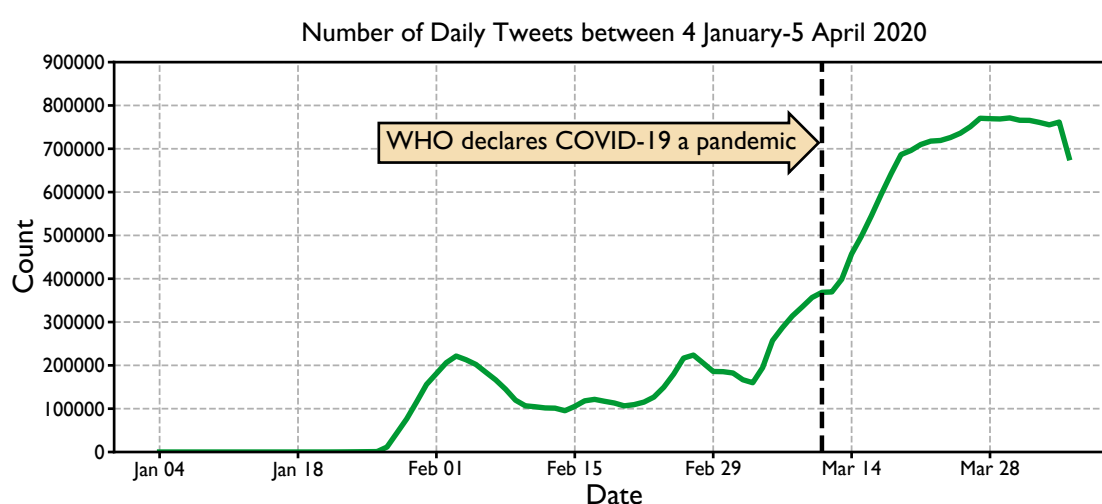


Figure 1. Daily Twitter activity related to coronavirus disease 2019 (COVID-19) during the early stages of the pandemic.

For training machine learning classifiers, we utilize the following two recently-curated datasets: *COVID-19 Intent* [75] and *COVID-19 Questions* [76]. Intent dataset consists of 4938

COVID-19 specific utterances (typically a question or a request) categorized into 16 categories to describe the author’s intent [75]. For instance, the sample “*is coughing a sign of the virus*” has an intent related to Symptoms. The dataset consists of English, French, and Spanish utterances and has been synthetically created by native-speaker annotators based on an ontology. We discard the uninformative categories of Hi and Okay/Thanks to end up with 4325 samples from this dataset. We combine Can_i_get_from_feces_animal_pets, Can_i_get_from_packages_surfaces, and How_does_corona_spread categories into a single category of Transmission. Similarly, we merge What_if_i_visited_high_risk_area category into Travel category to end up with 11 categories (classes).

Questions dataset consists of 1245 questions categorized into 16 categories collected from 13 sources [76]. Seven of the sources are frequently asked questions (FAQ) websites of recognized organizations such as the Center for Disease Control (CDC) and 6 of them are crowd-based sources such as Google Search. We use 594 samples from this dataset belonging to Prevention, Reporting, Speculation, Symptoms, Transmission, and Treatment categories. In the end, the dataset for our experiments, i.e., training and validating text classification algorithms, consists of 4,919 textual samples collected from the abovementioned two datasets. 11 category labels of the final dataset are Donate, News & Press, Prevention, Reporting, Share, Speculation, Symptoms, Transmission, Travel, Treatment, What Is Corona?. Sample distribution of languages and categories among the dataset can be examined from Tables 1 and 2, respectively.

Table 1. Distribution of languages.

Language	Samples
English	2119
French	1400
Spanish	1400
Total	4919

Table 2. Distribution of category labels.

Category	Samples
Donate	310
News & Press	310
Prevention	431
Reporting	389
Share	310
Speculation	363
Symptoms	348
Transmission	1152
Travel	615
Treatment	381
What Is Corona?	310
Total	4919

3.2. Tweet Embeddings

As the daily volume of COVID-19 related discussions on Twitter is enormous, computational public attention surveillance would benefit from lightweight approaches that can still maintain a high predictive power. Preferably, numerical representations should encode the semantics of tweets in such a way that simple vector operations should suffice for large-scale retrieval or even classification. Moreover, developed machine learning systems should be able to accommodate tweets in several languages to be able to capture the public discourse in an unbiased manner. Multilingual BERT-like contextual word/token embeddings [59] have been shown to be effective as pre-trained models if followed by a task-specific fine-tuning. However, they do not intrinsically produce effective

sentence-level representations [68]. In order to be able to take advantage of multilingual BERT encoders for extracting out-of-the-box sentence embeddings, we employ Language-agnostic BERT Sentence Embeddings [22].

LaBSE embeddings combine BERT-based dual-encoder framework with masked language modeling (an unsupervised fill-in-the-blank task where a model tries to predict a masked word) to reach state-of-the-art performance in embedding sentences across 109 languages [22]. Trained on a corpus of 6 billion translation pairs, LaBSE embeddings provide out-of-the-box comparison ability of sentences even by a simple *dot product* (essentially corresponding to *cosine* similarity as embeddings are l_2 normalized). We encode both the training data and 26.8 million tweets using this deep learning approach, ending up with vectors of length 768 for each observation. Embeddings are extracted with *TensorFlow* (version 2.2) framework in Python 3.7 on a 64 bit Linux machine with an NVIDIA Titan Xp GPU.

3.3. Intent Classification

As our choice of embeddings provide effective, out-of-the-box latent space representations of the textual data, simpler classifiers can be directly employed for identifying the prevalent topic of a tweet. In fact, LaBSE embeddings provide representations that are suitable to be compared with simple *cosine similarity* [22]. We train 3 classifiers (multi-class, single-label classification), namely k-nearest neighbor (kNN), logistic regression (LR), and support vector machine (SVM) to classify the observations into 11 categories. We employ a 10-fold cross-validation scheme to evaluate the performance of the three models. For comparison, we run the same experiments for multilingual BERT embeddings (base, uncased) as well. Hyperparameters of the classifiers are selected by Bayesian optimization (see Section 3.4). Once the embedding-classifier pair with its set of hyperparameters giving the highest cross-validation classification performance is selected, the classifier is trained with full dataset of 4919 observations. With this model, inference on 26,759,164 samples of Twitter data embeddings is performed.

3.4. Bayesian Hyperparameter Optimization

Typically, machine learning algorithms have several hyperparameters that require tuning for the specific task to avoid sub-optimal predictive performance. The most influential hyperparameters of k-nearest neighbor classifier are k (number of neighbors) and distance metric (e.g., *cosine* (Although, not an official distance metric as it violates triangle inequality.), *euclidean*, *manhattan*, etc.). For support vector machine classifier, trade-off between training error and margin (essentially regularization), C , and the choice of kernel function (linear, polynomial, or radial basis function) are the most crucial hyperparameters. l_2 regularization coefficient is the main hyperparameter for logistic regression classifier. We formulate the problem of finding the optimal set of classifier hyperparameters, $\hat{\theta}$, as a Bayesian optimization problem:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} f(\theta), \quad (1)$$

where $f(\theta)$ is the average of cross-validation accuracies for a given set of hyperparameters, i.e., $\frac{1}{N} \sum_{i=1}^N ACC_i$. For our experiments, $N = 10$ as we perform 10-fold cross-validation. We use *Gaussian Processes* for the surrogate model [77] of the Bayesian optimization by which we emulate the statistical relationships between the hyperparameters and model performance, given a dataset. We run the optimization scheme for 30 iterations (each iteration corresponds to one full cross-validation) for each classifier separately.

Bayesian optimization is especially beneficial in settings where the function to be minimized/maximized, $f(\theta)$, is a black-box function without a known closed-form and expensive to evaluate [78]. As $f(\theta)$ corresponds to cross-validation performance in our case, it indeed is a black-box function that is computationally expensive to evaluate. That was our motive for employing Bayesian

hyperparameter optimization instead of manual tuning or performing grid-search over a manually selected hyperparameter space.

3.5. Evaluation

For visual inspection of embeddings, we utilize Uniform Manifold Approximation and Projection (UMAP) to map the high dimensional embeddings (768 for LaBSE and 1024 for BERT) to a 2-dimensional plane [79]. UMAP is a frequently used dimensionality reduction and visualization technique that can preserve global structure of the data better than other similar methods [79]. In their recent study, Ordun et al. employ UMAP visualization of COVID-19 tweets as well [32].

Evaluation of classifiers and their sets of hyperparameters are performed by 10-fold cross-validation. Randomness (seed) in cross-validation splits are fixed in order to perform fair comparison. Average accuracy (%) and F1 scores (micro and macro averages) across 10 folds are reported for all classifiers (for their best performing set of hyperparameters) for BERT and LaBSE embeddings. Confusion matrix for the best performing representation-classifier pair is reported as well. After running inference on Twitter data to classify 26.8 million tweets into 11 categories with the best performing classifier, we aggregate the overall distribution of Twitter chatter into percentages. We also show tweet examples from each predicted category.

4. Results

UMAP visualization of LaBSE embeddings of the training data is depicted in Figure 2. Most visibly distinctive clusters belong to categories *Donate*, *Share*, and *Travel*. In this study, a cumulative of 88 h of GPU computation was performed for extracting language-agnostic embeddings for the 26.8 million tweets which roughly corresponds to a carbon footprint of 9.5 kgCO₂eq (estimated by following [80]).

The 10-fold cross-validation results for the classifiers with the highest scoring set of hyperparameters are shown in Table 3. For all classifiers, LaBSE embeddings outperform multilingual BERT embeddings. The best hyperparameters for k-nearest neighbor classifier were found to be $k = 7$ and *cosine* distance for both embeddings. Optimal regularization coefficients for logistic regression was found to be 4.94×10^3 for LaBSE and 1.01×10^2 for BERT representations. For support vector machine, optimal C was found to be 5.07 and 3.51 for LaBSE and BERT representations, respectively. The best choice of kernel function was found to be radial basis function for both representations. The best performing classifier was found to be support vector machine classifier applied on LaBSE embeddings with 86.92% accuracy, 0.876 micro-F1 score, and 0.881 macro-F1 score. Confusion matrix of this classifier out of cross-validation predictions can be examined from Figure 3. In parallel to visual findings in Figure 2, *Donate*, *Share*, and *Travel* classes reach high accuracies of 97.1%, 98.1%, and 94.0%, respectively. Classifier has the highest error rate for the *Prevention* and *Speculation* classes, both staying below 80% accuracy. Our results show that more than 15% of samples belonging to *Speculation* category have been misclassified as *Transmission*.

Figure 4 depicts the timeline of normalized daily category distributions obtained by running inference on tweets posted between 26 January and 5 April 2020. *Transmission* and travel-related chatter as well as speculations (opinions on origin of COVID-19, myths, and conspiracies) show significance presence throughout the pandemic. *What Is Corona?*, i.e., questions and inquiries regarding what exactly COVID-19 is, shows a presence in the early stages of the pandemic but decreases through time, possibly due to gained scientific knowledge about the nature of the disease. On the contrary, prevalence of *Prevention* related tweets increase through time especially after the declaration of pandemic by the WHO on March 11. Similarly, chatter for *Donation* discussions are observed only starting from March. Timeline curves become smoother (less spiky) with increasing date as the percentage changes between consecutive days gets smaller. This is intuitive as the total number of tweets in January is several magnitudes lower than that of April and sudden percentage jumps in January can be attributed to only a handful of tweets. Finally, random samples of tweets and their predicted labels can be observed in Table 4.

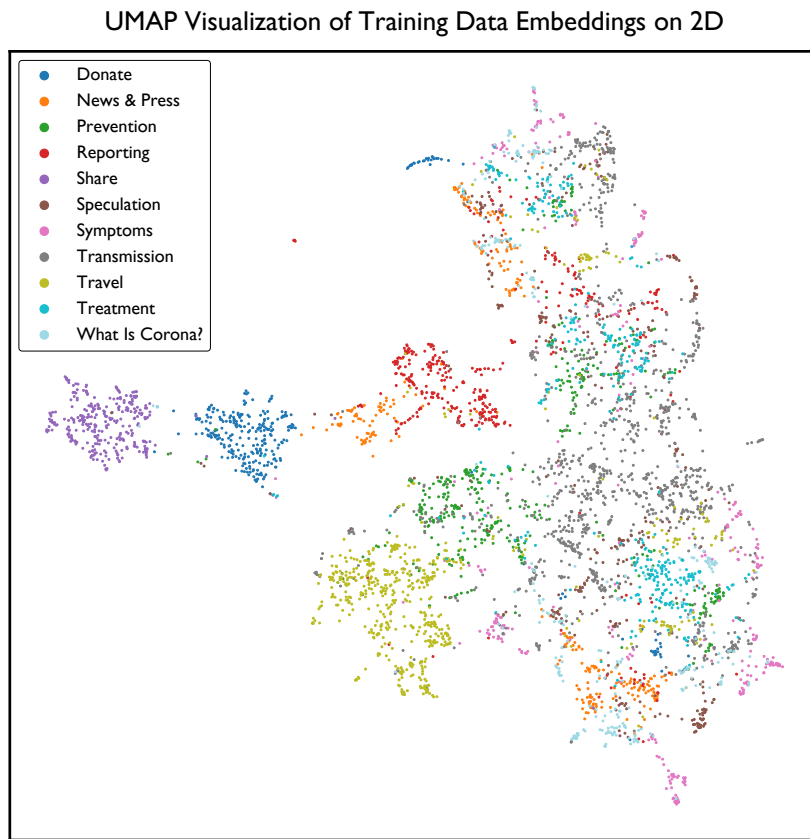


Figure 2. Uniform Manifold Approximation and Projection (UMAP) visualization of Language-agnostic BERT Sentence Embeddings belonging to 4919 observations among 11 COVID-19 discourse categories.

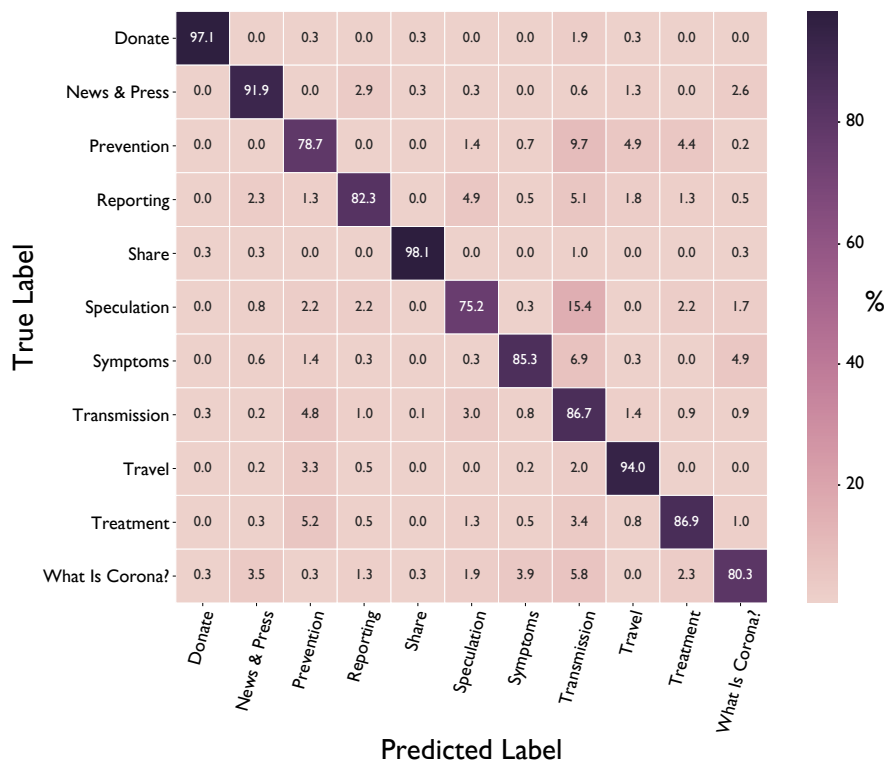


Figure 3. Normalized confusion matrix of SVM classifier predictions with Language-agnostic BERT Sentence Embeddings (LaBSE) across cross-validation folds.

Table 3. Cross-validation results of three classifiers for BERT and LaBSE embeddings.

Model	BERT			LaBSE		
	Accuracy (%)	F1 (Micro)	F1 (Macro)	Accuracy (%)	F1 (Micro)	F1 (Macro)
kNN	72.54	0.725	0.725	82.76	0.828	0.827
LR	76.62	0.766	0.771	86.05	0.844	0.846
SVM	81.81	0.818	0.820	86.92	0.876	0.881

Table 4. Example tweets and predicted classification categories.

Tweet	Predicted Class
China Providing Assistance To Pakistani Students Trapped in Wuhan: Ambassador—#Pakistan	Donate
Results are in. State health officials say three suspected cases of Coronavirus have tested NEGATIVE. There is a fourth possible case from Washtenaw County being sent to the CDC.	News & Press
what are good steps to protect ourselves from the Coronavirus?	Prevention
The first coronavirus case has been confirmed in the U.S. #virus	Reporting
Share this and save lives #coronavirus #SSOT	Share
#coronavirus Don't let these ignorant people make you believe that this corona virus is any different than SARS IN 2003 which was contained after a few months. They want you to panic as they have ulterior motives such as shorting the stock market etc.	Speculation
I have a rushing sound in my ears. It doesn't seem to match the symptoms for the #coronavirus so perhaps it is the sound of the #EU leaving my body...	Symptoms
what animals can carry Wuhan coronavirus?	Transmission
can we ban flights from wuhan pls!?!?	Travel
¿Qué medicamento nos colará en esta ocasión la industria farmacéutica para combatir al coronavirus?	Treatment
Oque é coronavirus?	What Is Corona?

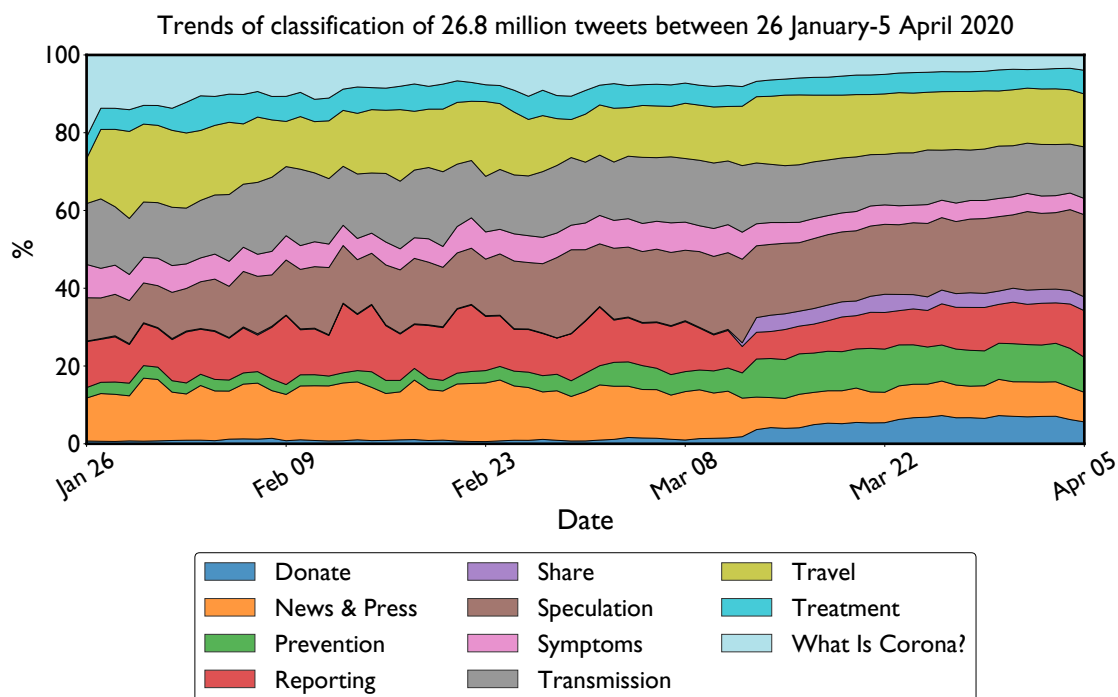


Figure 4. Distribution of semantic discussion categories in Twitter predicted by the classifier during COVID-19.

5. Discussion

Adequate risk management in crisis situations has to take into account not only the threat itself but also the perception of the threat by the public [81]. In the digital era, the public heavily relies on social media to inform their level of risk perception, often in a rapid manner. In fact, social media enhances collaborative problem-solving and citizens' ability to make sense of the situation during disasters [4]. With this paradigm in mind, we attempt to perform large-scale classification of 26.8 million COVID-19 tweets using natural language processing and machine learning. We utilize state-of-the-art language-agnostic tweet representations coupled with simple, lightweight classifiers to be able to capture COVID-19 related discourse during a span of 13 weeks.

Our first observation of "increasing Twitter activity with increased COVID-19 spread throughout the globe" (Figure 1) is in parallel with other studies. For instance, Bento et al. showed that Internet searches for "coronavirus" increased on the day immediately after the first case announcement for a location [82]. Wong et al. correlates announcement of new infections and Twitter activity [83]. Similar associations have been discovered between official cases and Twitter activity by causal modeling as well [69]. Secondly, we show that language-agnostic embeddings can be utilized in an out-of-the-box fashion, i.e., without requiring task-specific fine-tuning of BERT models. An SVM classifier reaches 86.92% accuracy and 0.881 macro-F1 score for classification into 11 topic categories. Finally, we show that overall public discourse shifts through the pandemic. Questions of "what coronavirus is" leave their place to donation and prevention related discussions as the disease spreads into more and more countries especially during March 2020. Tweets related to donation increase especially around 13 March 2020 when the WHO and the United Nations Foundation start a global COVID-19 donation fund [84].

When compared to existing studies that often employ unsupervised topic modeling, our approach tries to perform public attention surveillance with a more automated perspective as we formulate the problem as a supervised learning one. Topic modeling with LDA, which has been employed by majority of previous studies, relies on manual and qualitative inspection of discovered topics. Furthermore, plain LDA fails to accommodate contextual representations and does not assume a distance metric between discovered topics as it is based on the notion that words belonging to a topic are more likely to appear in the same document. With language-agnostic embeddings, we also include tweets from languages other than English to our analysis, hence decrease the selection bias.

Utilization of large-scale social media data for extracting health insights is even more pertinent during a global pandemic such as COVID-19, as running randomized control trials becomes less practical. Moreover, traditional surveys for public attention surveillance may further stress the participants whose mental health and overall well-being might have been affected by lockdowns, associated financial issues, and changes in social dynamics [85–87]. Once accurate estimation of global or national discourse is possible, social media can also be used to direct people to trusted resources, counteract misinformation, disseminate reliable information, and enable a culture of preparedness [88]. Considering the popularity of the social media platforms such as Twitter, it would be beneficial for these platforms to collaborate with expert epidemiologists to implement relevant digital tools that can provide fact-based insights to the public. Assessment of effectiveness of public risk communication and interventions is also feasible with these computational systems. Guided by machine learning insights, some of these interventions can be made on social media itself.

Our study has several limitations. First, the training data consist of single label annotations while in reality a tweet can have several topics simultaneously, e.g., Prevention and Travel. Secondly, we do not employ a confidence threshold for categorizing tweets which forces our model to classify every observation into one of the 11 categories. Considering some Twitter discourses related to COVID-19 may not be properly represented by our existing categories, a probability threshold can be introduced for the final classification decision. We believe formulating the problem as a quantification task instead of a classification task would benefit this study as we are mainly interested in aggregate-level

prevalence of topics [89–92]. Finally, we discard retweets in our analysis, which in fact contributes to public attention on Twitter.

Future research includes running similar analysis for a more granular category set or sub-categories. For instance, the *Speculation* category can be divided into conspiracies related to origin of the disease, transmission characteristics, and treatment options. Including up-to-date Twitter data (after April 2020) as well as extracting location-specific insights will be performed in future analyses as well.

6. Conclusions

Transforming social media data into actionable knowledge for public health systems faces several challenges such as advancing methodologies to extract relevant information for health services, creating dynamic knowledge bases that address disaster contexts, and expanding social media research to focus on health response activities [93]. We hope our study serves this purpose by proving methodologies for large-scale, language-agnostic discourse classification on Twitter.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BERT	Bidirectional Encoder Representations from Transformers
CDC	Center for Disease Control
COVID-19	Coronavirus disease 2019
FAQ	Frequently Asked Questions
kNN	k-nearest neighbor (kNN)
LaBSE	Language-agnostic BERT Sentence Embeddings
LDA	Latent Dirichlet Allocation
LR	Logistic regression
SVM	Support vector machine
UMAP	Uniform Manifold Approximation and Projection

References

1. Cucinotta, D.; Vanelli, M. WHO declares COVID-19 a pandemic. *Acta-Bio-Med. Atenei Parm.* **2020**, *91*, 157–160. [[CrossRef](#)]
2. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [[CrossRef](#)]
3. Mahase, E. Coronavirus: COVID-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* **2020**. [[CrossRef](#)] [[PubMed](#)]
4. Jurgens, M.; Helsloot, I. The effect of social media on the dynamics of (self) resilience during disasters: A literature review. *J. Conting. Crisis Manag.* **2018**, *26*, 79–88. [[CrossRef](#)]
5. Van Bavel, J.J.; Baicker, K.; Boggio, P.S.; Capraro, V.; Cichocka, A.; Cikara, M.; Crockett, M.J.; Crum, A.J.; Douglas, K.M.; Druckman, J.N.; et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **2020**, *4*, 460–471. [[CrossRef](#)]
6. Zhong, B.L.; Luo, W.; Li, H.M.; Zhang, Q.Q.; Liu, X.G.; Li, W.T.; Li, Y. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: A quick online cross-sectional survey. *Int. J. Biol. Sci.* **2020**, *16*, 1745. [[CrossRef](#)]
7. Signorini, A.; Segre, A.M.; Polgreen, P.M. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS ONE* **2011**, *6*, e0019467. [[CrossRef](#)]
8. Ji, X.; Chun, S.A.; Geller, J. Monitoring public health concerns using twitter sentiment classifications. In Proceedings of the 2013 IEEE International Conference on Healthcare Informatics, Philadelphia, PA, USA, 9–11 September 2013; pp. 335–344. [[CrossRef](#)]

9. Ji, X.; Chun, S.A.; Wei, Z.; Geller, J. Twitter sentiment classification for measuring public health concerns. *Soc. Netw. Anal. Min.* **2015**, *5*, 13. [[CrossRef](#)]
10. Weeg, C.; Schwartz, H.A.; Hill, S.; Merchant, R.M.; Arango, C.; Ungar, L. Using Twitter to measure public discussion of diseases: A case study. *JMIR Public Health Surveill.* **2015**, *1*, e6. [[CrossRef](#)]
11. Mollema, L.; Harmsen, I.A.; Broekhuizen, E.; Clijnk, R.; De Melker, H.; Paulussen, T.; Kok, G.; Ruiter, R.; Das, E. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. *J. Med. Internet Res.* **2015**, *17*, e128. [[CrossRef](#)]
12. Jordan, S.E.; Hovet, S.E.; Fung, I.C.H.; Liang, H.; Fu, K.W.; Tse, Z.T.H. Using Twitter for public health surveillance from monitoring and prediction to public response. *Data* **2019**, *4*, 6. [[CrossRef](#)]
13. Rosenberg, H.; Syed, S.; Rezaie, S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Can. J. Emerg. Med.* **2020**, *22*, 418–421. [[CrossRef](#)] [[PubMed](#)]
14. Chen, E.; Lerman, K.; Ferrara, E. COVID-19: The first public coronavirus Twitter dataset. *arXiv* **2020**, arXiv:2003.07372.
15. Gao, Z.; Yada, S.; Wakamiya, S.; Aramaki, E. NAIST COVID: Multilingual COVID-19 Twitter and Weibo Dataset. *arXiv* **2020**, arXiv:2004.08145.
16. Lamsal, R. Corona Virus (COVID-19) Tweets Dataset. *Sch. Comput. Syst. Sci.* **2020**. [[CrossRef](#)]
17. Aguilar-Gallegos, N.; Romero-García, L.E.; Martínez-González, E.G.; García-Sánchez, E.I.; Aguilar-Ávila, J. Dataset on dynamics of Coronavirus on Twitter. *Data Brief.* **2020**, *30*, 105684. [[CrossRef](#)]
18. Chen, E.; Lerman, K.; Ferrara, E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health Surveill.* **2020**, *6*, e19273. [[CrossRef](#)]
19. Abd-Alrazaq, A.; Alhuwail, D.; Househ, M.; Hamdi, M.; Shah, Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. *J. Med. Internet Res.* **2020**, *22*. [[CrossRef](#)]
20. Rao, H.R.; Vemprala, N.; Akello, P.; Valecha, R. Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. *Int. J. Inf. Manag.* **2020**, *55*, 102187. [[CrossRef](#)]
21. Park, H.W.; Park, S.; Chong, M. Conversations and medical news frames on twitter: Infodemiological study on covid-19 in south korea. *J. Med. Internet Res.* **2020**, *22*, e18897. [[CrossRef](#)]
22. Feng, F.; Yang, Y.; Cer, D.; Arivazhagan, N.; Wang, W. Language-agnostic BERT Sentence Embedding. *arXiv* **2020**, arXiv:2007.01852.
23. Dewhurst, D.R.; Alshaabi, T.; Arnold, M.V.; Minot, J.R.; Danforth, C.M.; Dodds, P.S. Divergent modes of online collective attention to the COVID-19 pandemic are associated with future caseload variance. *arXiv* **2020**, arXiv:2004.03516.
24. Thelwall, M.; Thelwall, S. Retweeting for COVID-19: Consensus building, information sharing, dissent, and lockdown life. *arXiv* **2020**, arXiv:2004.02793.
25. Alshaabi, T.; Minot, J.R.; Arnold, M.V.; Adams, J.L.; Dewhurst, D.R.; Reagan, A.J.; Muhamad, R.; Danforth, C.M.; Dodds, P.S. How the world's collective attention is being paid to a pandemic: COVID-19 related 1-gram time series for 24 languages on Twitter. *arXiv* **2020**, arXiv:2003.12614.
26. Hamamsy, T.C.; Bonneau, R. Twitter activity about treatments during the COVID-19 pandemic: Case studies of remdesivir, hydroxychloroquine, and convalescent plasma. *medRxiv* **2020**. [[CrossRef](#)]
27. Singh, L.; Bansal, S.; Bode, L.; Budak, C.; Chi, G.; Kawintiranon, K.; Padden, C.; Vanarsdall, R.; Vraga, E.; Wang, Y. A first look at COVID-19 information and misinformation sharing on Twitter. *arXiv* **2020**, arXiv:2003.13907.
28. Lopez, C.E.; Vasu, M.; Gallemore, C. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. *arXiv* **2020**, arXiv:2003.10359.
29. Kouzy, R.; Abi Jaoude, J.; Kraitem, A.; El Alam, M.B.; Karam, B.; Adib, E.; Zarka, J.; Traboulsi, C.; Akl, E.W.; Baddour, K. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus* **2020**, *12*. [[CrossRef](#)] [[PubMed](#)]
30. Wicke, P.; Bolognesi, M.M. Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PLoS ONE* **2020**, *15*, e0240010. [[CrossRef](#)]
31. Jarynowski, A.; Wójta-Kempa, M.; Belik, V. Trends in Perception of COVID-19 in Polish Internet. *medRxiv* **2020**. [[CrossRef](#)]

32. Ordun, C.; Purushotham, S.; Raff, E. Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs. *arXiv* **2020**, arXiv:2005.03082.
33. Medford, R.J.; Saleh, S.N.; Sumarsono, A.; Perl, T.M.; Lehmann, C.U. An “Infodemic”: Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the Coronavirus Disease 2019 Outbreak. *Open Forum Infect. Dis.* **2020**, *7*. [[CrossRef](#)] [[PubMed](#)]
34. Chen, L.; Lyu, H.; Yang, T.; Wang, Y.; Luo, J. In the eyes of the beholder: Sentiment and topic analyses on social media use of neutral and controversial terms for COVID-19. *arXiv* **2020**, arXiv:2004.10225.
35. Cinelli, M.; Quattrocioni, W.; Galeazzi, A.; Valensise, C.M.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The COVID-19 Social Media Infodemic. *Sci. Rep.* **2020**, *10*, 16598. [[CrossRef](#)] [[PubMed](#)]
36. Hosseini, P.; Hosseini, P.; Broniatowski, D.A. Content analysis of Persian/Farsi Tweets during COVID-19 pandemic in Iran using NLP. *arXiv* **2020**, arXiv:2005.08400.
37. Jang, H.; Rempel, E.; Carenini, G.; Janjua, N. Exploratory Analysis of COVID-19 Related Tweets in North America to Inform Public Health Institutes. *arXiv* **2020**, arXiv:2007.02452.
38. Saad, M.; Hassan, M.; Zaffar, F. Towards Characterizing the COVID-19 Awareness on Twitter. *arXiv* **2020**, arXiv:2005.08379.
39. Odlum, M.; Cho, H.; Broadwell, P.; Davis, N.; Patrao, M.; Schauer, D.; Bales, M.E.; Alcantara, C.; Yoon, S. Application of Topic Modeling to Tweets as the Foundation for Health Disparity Research for COVID-19. *Stud. Health Technol. Inform.* **2020**, *272*, 24–27. [[CrossRef](#)]
40. Park, S.; Han, S.; Kim, J.; Molaie, M.M.; Vu, H.D.; Singh, K.; Han, J.; Lee, W.; Cha, M. Risk Communication in Asian Countries: COVID-19 Discourse on Twitter. *JMIR* **2020**. [[CrossRef](#)]
41. Xue, J.; Chen, J.; Hu, R.; Chen, C.; Zheng, C.; Zhu, T. Twitter discussions and concerns about COVID-19 pandemic: Twitter data analysis using a machine learning approach. *JMIR* **2020**. [[CrossRef](#)]
42. Gupta, R.K.; Vishwanath, A.; Yang, Y. COVID-19 Twitter Dataset with Latent Topics, Sentiments and Emotions Attributes. *arXiv* **2020**, arXiv:2007.06954.
43. Wang, X.; Zou, C.; Xie, Z.; Li, D. Public Opinions towards COVID-19 in California and New York on Twitter. *medRxiv* **2020**. [[CrossRef](#)]
44. Feng, Y.; Zhou, W. Is Working From Home The New Norm? An Observational Study Based on a Large Geo-tagged COVID-19 Twitter Dataset. *arXiv* **2020**, arXiv:2006.08581.
45. Yin, H.; Yang, S.; Li, J. Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media. *arXiv* **2020**, arXiv:2007.02304.
46. McQuillan, L.; McAweeney, E.; Bargar, A.; Ruch, A. Cultural Convergence: Insights into the behavior of misinformation networks on Twitter. *arXiv* **2020**, arXiv:2007.03443.
47. Omoya, Y.; Kaigo, M. Suspicion Begets Idle Fears—An Analysis of COVID-19 Related Topics in Japanese Media and Twitter. *SSRN* **2020**. [[CrossRef](#)]
48. Sharma, K.; Seo, S.; Meng, C.; Rambhatla, S.; Dua, A.; Liu, Y. Coronavirus on Social Media: Analyzing Misinformation in Twitter Conversations. *arXiv* **2020**, arXiv:2003.12309.
49. Kabir, M.; Madria, S. Corona Vis: A Real-time COVID-19 Tweets Analyzer. *arXiv* **2020**, arXiv:2004.13932.
50. Rosa, K.D.; Shah, R.; Lin, B.; Gershman, A.; Frederking, R. Topical clustering of tweets. *SWSM* **2011**, *63*. Available online: http://www.cs.cmu.edu/~encore/sigir_swsm2011.pdf (accessed on 30 July 2020).
51. Kaleel, S.B.; Abhari, A. Cluster-discovery of Twitter messages for event detection and trending. *J. Comput. Sci.* **2015**, *6*, 47–57. [[CrossRef](#)]
52. Lo, S.L.; Chiong, R.; Cornforth, D. An unsupervised multilingual approach for online social media topic identification. *Expert Syst. Appl.* **2017**, *81*, 282–298. [[CrossRef](#)]
53. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014.
54. Pennington, J.; Socher, R.; Manning, C. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543. [[CrossRef](#)]
55. Vosoughi, S.; Vijayaraghavan, P.; Roy, D. Tweet2vec: Learning tweet embeddings using character-level cnn-lstm encoder-decoder. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, 17–21 July 2016; pp. 1041–1044. [[CrossRef](#)]

56. Dhingra, B.; Zhou, Z.; Fitzpatrick, D.; Muehl, M.; Cohen, W. Tweet2Vec: Character-Based Distributed Representations for Social Media. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, Canada, 30 July–4 August 2016; pp. 269–274. [CrossRef]
57. Liu, J.; He, Z.; Huang, Y. Hashtag2Vec: Learning hashtag representation with relational hierarchical embedding model. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 3456–3462. [CrossRef]
58. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. *arXiv* **2018**, arXiv:1802.05365.
59. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2019; pp. 4171–4186. [CrossRef]
60. Gencoglu, O. Deep Representation Learning for Clustering of Health Tweets. *arXiv* **2018**, arXiv:1901.00439.
61. Zhu, J.; Tian, Z.; Kübler, S. UM-IU@LING at SemEval-2019 Task 6: Identifying Offensive Tweets Using BERT and SVMs. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; pp. 788–795. [CrossRef]
62. Ray Chowdhury, J.; Caragea, C.; Caragea, D. Keyphrase extraction from disaster-related tweets. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 1555–1566. [CrossRef]
63. Chowdhury, J.R.; Caragea, C.; Caragea, D. On Identifying Hashtags in Disaster Twitter Data. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 498–506. [CrossRef]
64. Roitero, K.; Cristian, B.; Mea, V.D.; Mizzaro, S.; Serra, G. Twitter Goes to the Doctor: Detecting Medical Tweets Using Machine Learning and BERT. In Proceedings of the International Workshop on Semantic Indexing and Information Retrieval for Health from Heterogeneous Content Types and Languages, Lisbon, Portugal, 14–17 April 2020; Volume 2619.
65. Mazoyer, B.; Cagé, J.; Hervé, N.; Hudelot, C. A french corpus for event detection on twitter. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 6220–6227.
66. Nguyen, D.Q.; Vu, T.; Nguyen, A.T. BERTweet: A pre-trained language model for English Tweets. *arXiv* **2020**, arXiv:2005.10200.
67. Müller, M.; Salathé, M.; Kummervold, P.E. COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter. *arXiv* **2020**, arXiv:2005.07503.
68. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3982–3992. [CrossRef]
69. Gencoglu, O.; Gruber, M. Causal Modeling of Twitter Activity during COVID-19. *Computation* **2020**, *8*, 85. [CrossRef]
70. Baly, R.; Karadzhov, G.; An, J.; Kwak, H.; Dinkov, Y.; Ali, A.; Glass, J.; Nakov, P. What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Seattle, WA, USA, 5–10 July 2020. [CrossRef]
71. Kim, H.; Walker, D. Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media. *Harv. Kennedy Sch. Misinf. Rev.* **2020**, *1*. [CrossRef]
72. Gencoglu, O. Cyberbullying Detection with Fairness Constraints. *IEEE Internet Comput.* **2020**. [CrossRef]
73. Banda, J.M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Chowell, G. A Twitter Dataset of 150+ million tweets related to COVID-19 for open research. *Zenodo* **2020**. [CrossRef]
74. Covid-19 Twitter Chatter Dataset for Scientific Use. Available online: <http://www.panacealab.org/covid19/> (accessed on 30 July 2020).
75. Arora, A.; Shrivastava, A.; Mohit, M.; Lecanda, L.S.M.; Aly, A. Cross-Lingual Transfer Learning for Intent Detection of Covid-19 Utterances. 2020. Available online: <https://openreview.net/pdf?id=vP-CQG-ap-R> (accessed on 29 November 2020).

76. Wei, J.; Huang, C.; Vosoughi, S.; Wei, J. What Are People Asking About COVID-19? A Question Classification Dataset. In Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020, Seattle, WA, USA, 19 July 2020.
77. Rasmussen, C.E. Gaussian Processes in Machine Learning. In *Summer School on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71. [[CrossRef](#)]
78. Moćkus, J. On Bayesian methods for seeking the extremum. *Optimization Techniques IFIP Technical Conference*; Springer: Berlin/Heidelberg, Germany, 1975; pp. 400–404.55. [[CrossRef](#)]
79. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [[CrossRef](#)]
80. Lacoste, A.; Luccioni, A.; Schmidt, V.; Dandres, T. Quantifying the Carbon Emissions of Machine Learning. *arXiv* **2019**, arXiv:1910.09700.
81. Sandman, P.M. *Responding to Community Outrage: Strategies for Effective Risk Communication*; AIHA: Fairfax County, VA, USA, 1993.
82. Bento, A.I.; Nguyen, T.; Wing, C.; Lozano-Rojas, F.; Ahn, Y.Y.; Simon, K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11220–11222. [[CrossRef](#)] [[PubMed](#)]
83. Wong, C.M.L.; Jensen, O. The paradox of trust: Perceived risk and public compliance during the COVID-19 pandemic in Singapore. *J. Risk Res.* **2020**. [[CrossRef](#)]
84. COVID-19 Solidarity Response Fund. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/donate> (accessed on 30 July 2020).
85. Wang, C.; Pan, R.; Wan, X.; Tan, Y.; Xu, L.; Ho, C.S.; Ho, R.C. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1729. [[CrossRef](#)]
86. Cullen, W.; Gulati, G.; Kelly, B. Mental health in the Covid-19 pandemic. *QJM Int. J. Med.* **2020**, *113*, 311–312. [[CrossRef](#)]
87. Brooks, S.K.; Webster, R.K.; Smith, L.E.; Woodland, L.; Wessely, S.; Greenberg, N.; Rubin, G.J. The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *Lancet* **2020**, *395*, 912–920. [[CrossRef](#)]
88. Merchant, R.M.; Lurie, N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. *J. Am. Med. Assoc.* **2020**, 323. [[CrossRef](#)]
89. Forman, G. Counting Positives Accurately Despite Inaccurate Classification. In *European Conference on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 564–575.55. [[CrossRef](#)]
90. Forman, G. Quantifying Counts and Costs via Classification. *Data Min. Knowl. Discov.* **2008**, *17*, 164–206. [[CrossRef](#)]
91. Bella, A.; Ferri, C.; Hernández-Orallo, J.; Ramirez-Quintana, M.J. Quantification via Probability Estimators. In Proceedings of the IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 737–742. [[CrossRef](#)]
92. González, P.; Díez, J.; Chawla, N.; del Coz, J.J. Why Is Quantification an Interesting Learning Problem? *Prog. Artif. Intell.* **2017**, *6*, 53–58. [[CrossRef](#)]
93. Chan, J.L.; Purohit, H. Challenges to transforming unconventional social media data into actionable knowledge for public health systems during disasters. *Disaster Med. Public Health Prep.* **2019**, doi:10.1017/dmp.2019.92. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).