Dr.Emotion: Disentangled Representation Learning for Emotion Analysis on Social Media to Improve Community Resilience in the COVID-19 Era and Beyond

Anonymous Author(s)

ABSTRACT

The deadly outbreak of coronavirus disease (COVID-19) has posed grand challenges to human society; in response to the spread of COVID-19, many activities have moved online and social media has played an important role by enabling people to discuss their experiences and feelings of this global crisis. To help combat the prolonged pandemic that has exposed vulnerabilities impacting community resilience, in this paper, based on our established large-scale COVID-19 related social media data, we propose and develop an integrated framework (named Dr.Emotion) to learn disentangled representations of social media posts (i.e., tweets) for emotion analysis and thus to gain deep insights into public perceptions towards COVID-19. In Dr.Emotion, for given social media posts, we first post-train a transformer-based model to obtain the initial post embeddings. Since users may implicitly express their emotions in social media posts which could be highly entangled with the content in the context, to address this challenge for emotion analysis, we propose an adversarial disentangler by integrating emotion-independent (i.e., sentiment-neutral) priors of the posts generated by another post-trained transformer-based model to separate and disentangle the implicitly encoded emotions from the content in latent space for emotion classification *at the first attempt*. Extensive experimental studies are conducted to fully evaluate Dr.Emotion and promising results demonstrate its performance in emotion analysis by comparisons with state-of-the-art baseline methods. By exploiting our developed Dr.Emotion, we further perform emotion analysis based on a large number of social media posts (i.e., 107,434 COVID-19 related tweets posted by users in the United States through Mar 1, 2020 to Sep 30, 2020) and provide in-depth investigation from both temporal and geographical perspectives, based on which additional work can be conducted to extract and transform the constructive ideas, experiences and support into actionable information to improve community resilience in responses to a variety of crises created by COVID-19 and well beyond.

KEYWORDS

Disentangled Representation Learning, Emotion Analysis, Social Media, COVID-19, Community Resilience.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The fast evolving and deadly outbreak of COVID-19 [48] has posed unprecedented challenges to human society. The United States (U.S.) has been one of the hardest-hit countries, accounting for around 20% of global cases and deaths from the disease as of Oct 15, 2020 [24]. In response to the spread of COVID-19, many of the states issued stay-at-home orders in March/April and began reopening with caution since May. The COVID-19 crisis has become a paradigm shifting phenomenon affecting hundreds of millions of people in the U.S. directly or indirectly - i.e., as illustrated in Figure 1.(a), it has led to escalating societal, economic, and behavioral issues with significant consequences evidenced by: (1) *Dramatic rise in unemployment.* The number of unemployed persons in the U.S. were up to 12.6 million in Sep 2020 (i.e., 7.9% unemployment rate compared to 3.68% in 2019 before the pandemic) [6]. (2) *Significant economic downturn.* Due to the closure of many businesses (e.g., 43% of small businesses [4]), the real gross domestic product (GDP) decreased at an annual rate of 31.4% and profits from current production decreased $208.9 billion in the second quarter of 2020 [5]. (3) *A variety of crises induced by the pandemic.* The anxiety, social isolation, unemployment stress, grief, and general uncertainty have created a variety of crises [10], including an impact on the volume and distribution of crime, increasing substance abuse and suicides. For example, compared to last year, the pandemic has seen a surge in crimes like domestic violence up 30% in New York state [38] in April and a 120% increase in gun violence in Chicago since mid-March [42]; while opioid (e.g., fentanyl) overdose deaths increased 133% in April-June in King County at Washington state [26].

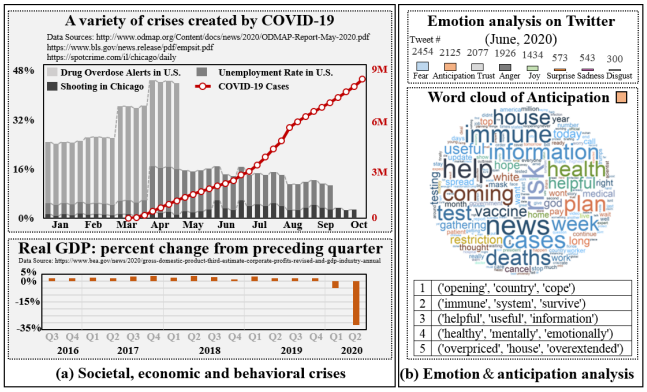


Figure 1: A variety of crises caused by COVID-19 and emotion analysis on Twitter in the United States.

The pandemic has exposed vulnerabilities impacting community resilience, including the inability to effectively and efficiently address the social, economic and behavioral issues, as expressed on social media (e.g., Twitter): a user experienced "*sad, depressed and*

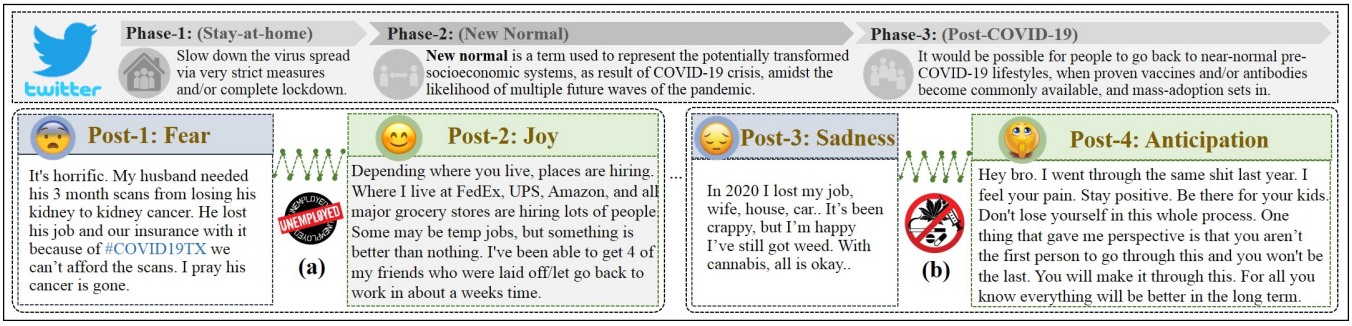


Figure 2: A path of emotion analysis on social media to improve community resilience in the COVID-19 era and beyond.

terrified" due to losing job and health insurance after working in a company for 17 years; an unemployed law school graduate felt "*so painful*"; a nurse in emergency room shared "*suicidal thoughts*" due to "*overwhelmed feeling*"; and another user posted "*bored of suffering, want drugs*". Even with reopening, as shown in Figure 1.(b), our analysis of 11,432 annotated tweets in the U.S. in June shows that the prolonged pandemic with a surge of COVID-19 cases since late June has caused dominant fear and volatile emotions in combination with anticipation and trust; the analysis of users' anticipation also demonstrates that one of the most anticipated demands is "*helpful and useful information*". How to derive such information to improve community resilience in responses to the devastating societal, economic, and behavioral effects caused by the pandemic?

To answer the above question, as social media has played a significant role by enabling people to discuss their experiences and feelings of COVID-19, in this paper, we propose to perform automatic emotion analysis of social media posts to i) assist with insights of what people are suffering and what helps they need, and ii) to identify experiences, ideas and support from the ones who are successfully navigating threats. For example, as shown in Figure 2, a Twitter user expressed her "*fear*" due to unemployment issue, while another user offered constructive suggestion with "*joy*" to help address the challenge: "*Where I live at FedEx, UPS, Amazon, USPS, and all major grocery stores are recruiting and hiring lots of people*"; and some users encouraged people with "*anticipation*" (e.g., "*Stay positive. Be there for your kids*", "*Call and talk with your friends*") to help people who may be with feelings of "*sadness*". As initial efforts, based on social media data (i.e., using Twitter as showcase in this work), we will focus on automatic emotion analysis that seeks to extract finer-grained emotions rather than coarse-grained polar sentiments to provide insights into public feelings, based on which additional work can be conducted to extract and transform the constructive ideas, experiences and support into actionable information to improve community resilience in responses to a variety of crises created by COVID-19 and well beyond.

To automate the emotion analysis on social media, there have been ample research studies: lexicons [37] and traditional machine learning models (e.g., naive Bayes [18], support vector machine (SVM) [50], logistic regression [35]) based on n-grams have been extensively explored; in recent years, deep learning models (e.g., convolutional neural networks (CNNs) [28, 54], recurrent neural networks (RNNs) [34, 47], Transformer [46]) have been developed for emotion/sentiment analysis; in particular, the pre-trained transformer based methods (e.g., bidirectional encoder representations from transformers (BERT) [12], robustly optimized BERT approach (RoBERTa) [32]) have achieved state-of-the-art performance. To perform emotion analysis for the insights of public feelings related to COVID-19, as shown in Figure 2, we observe that users may implicitly express their emotions in social media posts which are highly entangled with other descriptive information; more specifically, compared with the other three posts (*Post-1*, *Post-3* and *Post-4*), the emotion is more implicitly encoded across *Post-2*. This poses a new challenge for the pre-trained transformer based models (e.g., *Post-2* is classified as "*sadness*" by both BERT and RoBERTa), since they are task-agnostic (i.e., they are trained for general tasks like next sentence prediction, masked language modeling which are not directly related to the downstream tasks such as emotion analysis).

How to derive disentangled representations in latent space that could separate emotions from the content in social media posts for emotion analysis? Recently, disentangled representation learning has shown its success in computer vision [16, 21]; it has also drawn increasing attentions in natural language processing (NLP) [3, 23, 25, 31], most of which mainly focus on text generation by either preserving text content while reducing stylistic properties [23, 25, 31] or disentangling syntax and semantics [3, 7]. Different from these works, we aim to disentangle the encoded emotions from the content in social media posts. To achieve this goal, in this paper, we propose and develop an integrated framework named **Dr.Emotion** (shown in Figure 3) to learn disentangled representations of social media posts for emotion analysis and thus to gain insights into public feelings of COVID-19. In Dr.Emotion, (1) as illustrated in Figure 3.(a), for each given social media post (i.e., tweet), we first obtain its post embedding by post-training a transformer-based model (i.e., RoBERTa shown in Figure 3.(b)); (2) to learn disentangled representations of given posts for emotion analysis, we then propose to integrate emotion-independent (i.e., sentiment-neutral) priors generated by another post-trained transformer-based model (shown in Figure 3.(c)) to devise an adversarial disentangler (Figure 3.(d)) to separate and disentangle the embedded emotions from the content in latent space; afterwards, (3) the learned disentangled representations will be fed to a deep neural network (DNN) with three-layer Multilayer Perceptron (MLP) to train the classifier for emotion analysis to access public feelings of COVID-19 that can thus help improve community resilience. The major contributions of our work are summarized below.

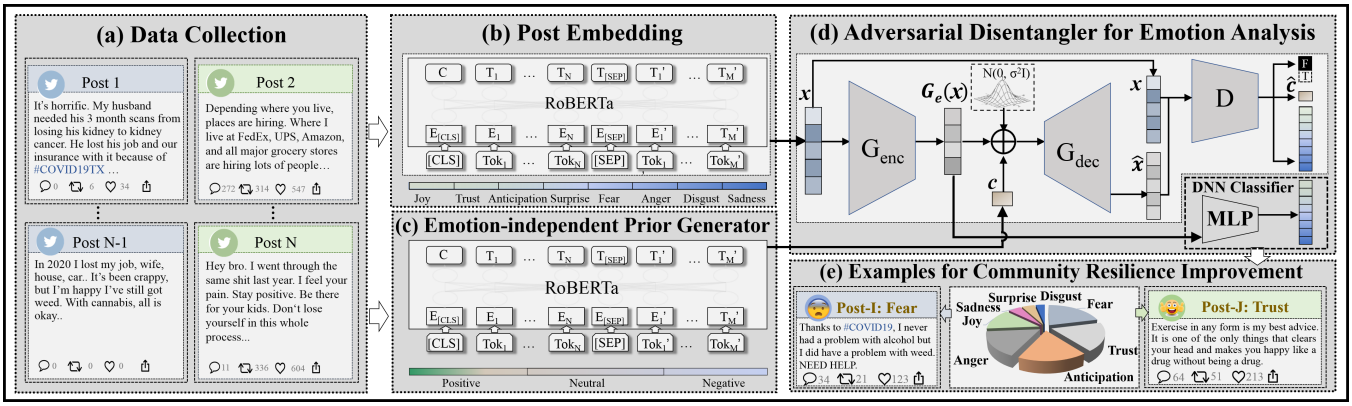


Figure 3: The overview of our proposed framework DR.Emotion.

- *We propose a novel adversarial disentangler for emotion analysis of social media posts.* As a user may implicitly express his/her emotion in social media post which could be highly entangled with the content, to address this challenge for emotion analysis, different from existing works, we propose an adversarial disentangler by integrating emotion-independent prior of the post generated by a post-trained transformer-based model to separate and disentangle the implicitly encoded emotion from the content in latent space *at the first attempt*. The proposed learning paradigm is a general framework to identify and disentangle the latent explanatory factors hidden in social media post data and thus can be applied to various web mining tasks.
- *We perform comprehensive experiments for model evaluation.* In this work, we have collected large-scale COVID-19 related tweets (142,434 tweets from Mar 1, 2020 to Sep 30, 2020) for emotion analysis; we have also built a benchmark dataset, consisting of 27,999 tweets manually labeled as one of the eight classes: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. Based on the constructed benchmark dataset, we perform comprehensive experimental studies to fully evaluate the performance of Dr.Emotion. Promising results demonstrate its efficacy in emotion analysis by comparisons with state-of-the-art methods.
- *Our proposed learning paradigm and extensive investigation based on the classified emotions will provide in-depth insights and customized guidance that can help public health experts, social workers and policy makers in decision-making and also enable a conceptual framework for the development of resilient community engagement strategies in responses to a variety of crises created by COVID-19 and well beyond.* To gain in-depth insights into public feelings of COVID-19, based on our large-scale social media data (i.e., 107,434 COVID-19 related tweets posted by users in the U.S. through Mar to Sep), we further deploy our developed Dr.Emotion in the wild to analyze how emotions change over time and across states in country. With the findings derived based on the classified emotions and by acknowledging what people truly anticipate or fear (trust, anger, etc.), actions could be implemented to mitigate negative ripples by policy makers, public health experts, business owners, organizations (e.g., schools), and any individual of interest.

2 PROPOSED METHOD

In this section, we introduce our proposed method of **d**isentangled **r**epresentation learning for **emotion** analysis (i.e., **Dr.Emotion**) in detail, which can help improve community resilience in the COVID-19 era and beyond.

2.1 COVID-19 Related Post Collection

To perform emotion analysis on social media (i.e., in this work, we initialize our efforts with the focus on Twitter) for insights of public feelings related to COVID-19, we have developed a set of web crawling tools using the Twitter search API [44] to collect tweets including COVID-19 related keywords or hashtag (e.g., "*COVID-19*", "*coronavirus*", "*corona*", "*pandemic*", etc.) on a daily basis from Mar 1, 2020 to Sep 30, 2020; and then we preprocess all tweets by converting all characters to lowercase and removing any retweet flag "RT", hashtag, mention indicator "@", emoji and URL. We further limit the locations of those tweets within the U.S. by checking location information in user profiles (note that all users are anonymized in our work using hash values of usernames) and thus we have had 142,434 tweets. To obtain the ground-truth for further investigation, two groups of annotators (three annotators per group) proficient in English have spent 30 days to manually label 35,000 randomly selected tweets into the following eight classes:

- **Anger**: e.g., "*There's way too many people out there who think this corona virus shit is over.*"
- **Anticipation**: e.g., "*Please trust the science. When it comes to this virus, science has come a long way and we are urging you to cooperate. Take this seriously. People for the love of the planet goodness gracious.*"
- **Disgust**: e.g., "*This idiot does not know the difference between a virus and a bacteria. He has also claimed they will have a cure for corona virus. Someone tell 'stable genius' no one has been able to kill a virus that is why we still have the common cold.*"
- **Fear**: e.g., "*I'm so scared because I'm worried for my son. I don't care about me but I pray my family dont catch this awful illness. This is taking such a toll on my mental health and well being.*"
- **Joy**: e.g., "*I cannot wait to tell my great great grandkids how I survived the great coronavirus scare of 2020.*"

- **Sadness**: e.g., "*There are currently 237 positive cases of coronavirus that have been identified among brookline residents and 18 of those people have died as a result of COVID-19.*"
- **Surprise**: e.g., "*On Fox news suddenly a very different tune about the coronavirus.*"
- **Trust**: e.g., "*Agreed, and it's also the time for employees to look out for each other and the business. If done right, with sacrifice by all together, the team can make it through this tough time #coronavirus #Employment.*"

Only those with mutual agreement are retained (i.e., the ones with conflicted labels by different groups will be excluded). Thus, we finally obtain 27,999 tweets with class labels of emotions.

2.2 Post Embedding

Transformer based models (e.g., BERT [12], RoBERTa [32]) have demonstrated significant superiority over traditional neural networks in NLP tasks such as next sentence prediction and masked language modeling. Those models are pre-trained with a large amount of English literature and thus can be used as a plug-and-play module for different applications. The key mechanism behind transformer based models is attention [46]. Specifically, unlike RNNs where the dependencies between distant tokens are difficult to learn [22], the transformer is able to assess any token in just one step, which is controlled by a learnable weight matrix. Formally, the attention mechanism can be formulated as following:

$$Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

where Q is the embedding of a token in the text (e.g., given social media post), V is the matrix of embeddings for all remaining tokens, K is the transformation matrix of V used to calculate the relative association between Q and V, and d_k is the dimension of keys K. The transformer models are usually constructed with multi-head attention modules:

$$MultiHead(Q, K, V) = (\bigoplus_{i=1}^{h} Attention(QW_i^Q, KW_i^K, VW_i^V))W^O, \tag{2}$$

where $\bigoplus$ denotes the concatenation operation, h is the number of heads, W_i^Q, W_i^K, W_i^V and W^O are transformation matrices.

Due to the superior performance in NLP tasks, for each given post, we exploit RoBERTa to obtain initial embedding in this work. Despite the success, RoBERTa [32] is pre-trained using books corpus and English Wikipedia used in BERT with additional data of Common Crawl news dataset, web text corpus and stories from Common Crawl, leading to task-agnostic and little understanding of social media posts. Especially in the COVID-19 themed emotion analysis scenario, the lack of abundant labeled data limits the fine-tune procedure. To solve this problem, we propose RoBERTa post-training to obtain better post embeddings. Taking its pre-trained weights as the initialization, based on the labeled dataset described in Section 2.1, we post-train RoBERTa by connecting feed-forward, dropout and output layers to the first transformer module in the last layer of RoBERTa; this transformer outputs the embedding that aggregates the embeddings from all other transformer nodes. The parameters can be optimized in an end-to-end manner by minimizing the following objective:

$$\mathcal{L} = -(\sum_{n=1}^{N}\sum_{l=1}^{L} y_{n,l}\log(\hat{y}_{n,l})) + \|\Theta\|_2^2, \tag{3}$$

where N and L are the number of social media posts (i.e., tweets) and the number of classes (i.e., eight emotion classes) respectively; $\hat{y}_{n,l}$ is the predicted score of post n in class l, $y_{n,l}$ is the class label, $\|\Theta\|_2^2$ is the L2-regularizer to prevent over-fitting. After post-training RoBERTa, we extract the hidden embedding of the first transformer module as the post embedding for each given tweet.

2.3 Emotion-independent Prior Generation

Although the above post embeddings obtained by the post-trained RoBERTa can be directly exploited for emotion analysis, as previously discussed, users may implicitly express and distribute their emotions in social media posts (i.e., the emotions could be highly entangled with the content), and thus simply using post embeddings by the post-training operation above may inadvertently and adversely impact overall performance. To solve this problem, as external knowledge could make the learned disentangled representations more reliable [33], we propose to generate and incorporate emotion-independent (i.e., sentiment-neutral) priors to guide the learner to separate emotions from the content in latent space for emotion analysis. More specifically, to disentangle the emotions encoded in the content in post embeddings, we propose to first generate the emotion-independent priors of given posts as conditions by sentiment analysis [49] that is a task of identifying if an expression is neutral or polar (i.e., positive/negative). To do this, taking its pre-trained weights as the initialization and based on our constructed dataset of 3,000 tweets manually labeled as either neutral or polar, we post-train another RoBERTa by adding dense, dropout and output layers to the first transformer in the last layer of RoBERTa, and optimizing the parameters of the entire network. In the end, we retrieve the neutral value c in the output layer as the emotion-independent prior for each given post:

$$c = \frac{exp(o_{neutral})}{\sum_{j\in\{neutral,polar\}} exp(o_j)}, \tag{4}$$

where $o_{neutral}$ and o_{polar} indicate the output neurons of neutral and polar classes respectively. Using the above post-trained RoBERTa for sentiment analysis, the higher value of c indicates the given post is more emotion-independent (i.e., sentiment-neutral) and vice versa, which can be illustrated by following examples.

- Tweet-1 **(c=0.4522)**: "*Depending where you live, places are hiring. Where I live at FedEx, UPS, Amazon, and all major grocery stores are hiring lots of people. Some may be temp jobs, but something is better than nothing. I've been able to get 4 of my friends who were laid off/let go back to work in about a weeks time.*"
- Tweet-2 **(c=0.0364)**: "*I'm so scared because I'm worried for my son. I don't care about me but I pray my family dont catch this awful illness. This is taking such a toll on my mental health and well being.*"

2.4 Adversarial Disentangler

Given a social media post, based on the initial embedding $\mathbf{x}$ obtained from Section 2.2, to integrate its emotion-independent prior

c (i.e., the condition generated from Section 2.3) to separate the encoded emotion from the content in latent space, we propose an adversarial disentangler (as shown in Figure 4) to achieve this goal. The proposed adversarial disentangler consists of a generator G and a discriminator D: given a post embedding (i.e., $\mathbf{x}$), G aims to incorporate the related emotion-independent prior c and Gaussian noise $\mathbf{z}$ (i.e., $\mathbf{z} \sim \mathcal{N}(0, \sigma)$) to produce a synthetic embedding $\hat{\mathbf{x}}$ via an encoder-decoder framework; D competes with G while assuring the emotion classification performance and retaining the prior; and the disentangled representation $G_e(\mathbf{x})$ will be derived via the adversarial minimax game [19]. We introduce the generator and discriminator in our designed framework in detail below.

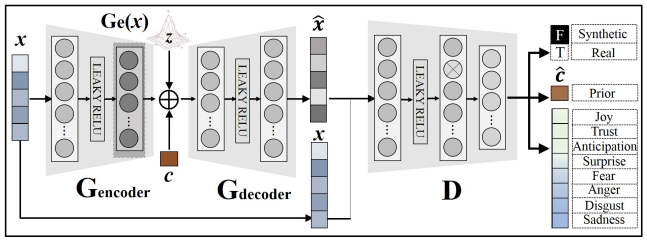


Figure 4: The framework of adversarial disentangler.

Multi-task Discriminator. The discriminator is a multi-task DNN consisting of three parts: $D = [D^T, D^L, D^C]$, where D^T aims to distinguishes a real post embedding from a synthetic one, D^L predicts the class label of the input (i.e., one in eight emotion classes) and D^C is to decode the related emotion-independent prior (i.e., $c \sim p(c)$). Given an input embedding, i.e., either a real post embedding $\mathbf{x}$ or a synthetic embedding $\hat{\mathbf{x}} = G(\mathbf{x}, c, \mathbf{z})$, the objective of D is to classify it as real or synthetic while estimating its emotion class label and decoding the related emotion-independent prior. The objective function of D is formulated as:

$$\begin{aligned} &\max_D V_D(D,G) = \\ &\mathbb{E}_{p_{\text{data}}(\mathbf{x})}[\log D^T(\mathbf{x}) + \log D^L(\mathbf{x}) - \log D^C(\mathbf{x})] + \\ &\mathbb{E}_{p_{\text{model}}(\mathbf{x})}[\log(1 - D^T(\hat{\mathbf{x}})) + \log D^L(\hat{\mathbf{x}}) - \log D^C(\hat{\mathbf{x}})]. \end{aligned} \tag{5}$$

The first part of Eq. (5) is to maximize the probabilities of real input $\mathbf{x}$ being identified as true and correctly classified while retaining the emotion-independent prior (i.e., we use the cross-entropy loss for D^T and D^L, and mean squared error for D^C); the second part conducts similar computations for class label and emotion-independent prior but attempts to maximize the probability of synthetic $\hat{\mathbf{x}}$ being identified as false.

Disentangled Generator. Given a post embedding $\mathbf{x}$, through an encoder-decoder framework, the generator G aims to generate a synthetic embedding $\hat{\mathbf{x}}$ with the incorporation of its related emotion-independent prior c. The generator is designed in the form of $G = [G_e, G_d]$, where the encoder G_e aims to learn a mapping from the original post embedding $\mathbf{x}$ to a disentangled representation $G_e(\mathbf{x})$, while the decoder G_d takes $G_e(\mathbf{x})$ together with the emotion-independent prior c and Gaussian noise $\mathbf{z}$ to produce $\hat{\mathbf{x}} = G_d(G_e(\mathbf{x}), c, \mathbf{z})$ aiming to fool D. The objective function of G can be formulated as:

$$\begin{aligned} \max_G V_G(D,G) = &\mathbb{E}_{\substack{p_{\text{data}}(\mathbf{x}) \\ p(\mathbf{c}),p(\mathbf{z})}}[\log(D^T(G_d(G_e(\mathbf{x}),c,\mathbf{z})))] \\ &+ \log(D^L(G_d(G_e(\mathbf{x}),c,\mathbf{z}))) \\ &- \log(D^C(G_d(G_e(\mathbf{x}),c,\mathbf{z})))]. \end{aligned} \tag{6}$$

Intuitively, our proposed adversarial disentangler exploits a "drop-and-recover" mechanism. While the post embedding $\mathbf{x}$ is passed through G_e with a reduced layer structure, a certain amount of information will be mandatorily dropped. By explicitly feeding the emotion-independent prior c with the output of G_e into G_d, c can be recovered in the synthetic embedding generated by G_d. Through the adversarial training process, G_e is expected to learn a mapping from original post embedding to the disentangled representation where c is reduced. The training process of the proposed *DR.Emotion* is given in Algorithm 1.

Algorithm 1: Training algorithm of Dr.Emotion

Input: post embeddings $\mathbf{X}$, emotion-independent priors $\mathbf{C}$, Gaussian noises $\mathbf{Z}$, emotion class labels L
Output: disentangled representation $G_e(\mathbf{x})$, $\mathbf{x} \in \mathbf{X}$

```
for each epoch do
    for each batch do
        /* Multi-task Discriminator Training */
        Sample a half batch of real inputs < x, c, l >, where
          x ∈ X, c ∈ C, l ∈ L;
        Update D by Eq. (5);
        Generate related half batch of synthetic inputs
          < G_d(G_e(x), c, z), c, l >;
        Update D by Eq. (5);
        Freeze D;
        /* Disentangled Generator Training */
        Generate a batch of synthetic inputs
          < G_d(G_e(x), c, z), c, l >;
        Update G by Eq. (6);
    end
end
return G_e(x);
```

2.5 Emotion Classification

To this end, for a social media post (i.e., tweet) with the initially obtained embedding $\mathbf{x}$, its disentangled representation $G_e(\mathbf{x})$ will be derived by the above proposed adversarial disentangler. We then directly feed the learned representation to train a three-layer MLP (as shown in Figure 3.(d)) for emotion classification task (i.e., classifying emotion of the post into one of the following eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise and trust). For complexity analysis, given N posts with d-dimensional embeddings as inputs, the time complexity of training the proposed adversarial disentangler in each epoch is $O(Nd^2 \prod h) \approx O(N)$, as h is the number of hidden neurons which is constant.

3 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we conduct extensive experimental studies using the collected large-scale COVID-19 related tweets to fully evaluate the performance of *DR.Emotion* for emotion analysis by comparisons with various state-of-the-art methods.

3.1 Experimental Setup

Data Collection and Annotation. As introduced in Section 2.1, based on our collected and preprocessed 142,434 COVID-19 related tweets, to obtain the ground-truth for further investigation, two groups of annotators (three annotators per group) proficient in English have spent 30 days to manually label 35,000 randomly selected tweets into the following eight classes: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. Only those with mutual agreement are retained (i.e., the ones with conflicted labels by different groups will be excluded); thus, we finally obtain 27,999 labeled tweets. Table 1 gives the summary of our collected and annotated dataset that will be used in the experiments.

Table 1: The collected/annotated COVID-19 related data.

Total	**142,434**	**Annotated**	**27,999**
Class	Tweet # (%)	Class	Tweet # (%)
Anger	4,467 (15.96%)	Anticipation	4,893 (17.48%)
Disgust	853 (3.05%)	Fear	5,815 (20.77%)
Joy	4,350 (15.54%)	Sadness	1,342 (4.79%)
Surprise	1,378 (4.92%)	Trust	5,011 (17.90%)

Environmental Settings. The experimental studies are conducted in the Ubuntu 18.04.4 LTS operating system, plus Ryzen 3900X, 2-way SLI GeForce RTX 2080 Ti and 64GB of RAM, with the framework of Python 3.7.6, TensorFlow 2.2.0 and HuggingFace 3.1.0.

Hyperparameters and Reproducibility. In our experiments, for the post-trained transformer based models (i.e., RoBERTa) in Section 2.2 and 2.3, we set the post embedding dimension to 768 and the number of epochs to 10; we perform Adaptive Moment Estimation (Adam) to optimize the models with a learning rate of 0.00003 and L2 regularization $\gamma = 10^{-4}$. For the adversarial disentangler in Section 2.4, we use a five-layer MLP as the encoder and decoder modules in generator and a six-layer MLP for the discriminator; and the encoder and decoder in generator are designed in a symmetric fashion. For DNN classifier, we use a three-layer MLP to build the model. We set the disentangled representation dimension to 128 and the number of epochs to 500; and we also use Adam for optimization with a learning rate of 0.0002. For other parameters, we set the batch size to 16 and the dropout rate to 0.2, and use LeakyReLU with the negative slope coefficient $\alpha = 0.2$ as the activation function. We also include our sample dataset and open-source codes for reproducibility, which are available in GitHub [1].

Evaluation Metrics. To quantitatively analyze the performance of different methods for multi-class emotion classification, based on our labeled dataset summarized in Table 1, we perform 10-fold cross validations and adopt the following commonly-used measures [45] for evaluation: Micro-AUC (Area Under the Curve), Macro-AUC, Micro-F1 and Macro-F1.

3.2 Evaluation of Dr.Emotion

In this set of experiments, we comprehensively validate our proposed methods integrated in *Dr.Emotion* for tweet emotion analysis, including post embedding and the adversarial disentangler.

In Dr.Emotion, we post-train RoBERTa [32] using our labeled dataset summarized in Table 1 to obtain initial post embeddings. Here, we compare it with following state-of-the-art transformer based models for this task (i.e., BERT [12], XLNet [52] and XLM [30]). Similar to RoBERTa in our framework, we also fine-tune these three transformer based models to generate the post embeddings.

- **BERT** [12] is pre-trained using next sentence prediction and masked language modeling with 800M book corpus words and 2,500M English Wikipedia words.
- **XLNet** [52] conducts permutation language modeling, two-stream self-attention and partial prediction for the pre-training process with 16GB texts.
- **XLM** [30] implements causal, masked, translation and cross-lingual language modeling, with MultiUN, IIT Bombay and OPUS datasets.

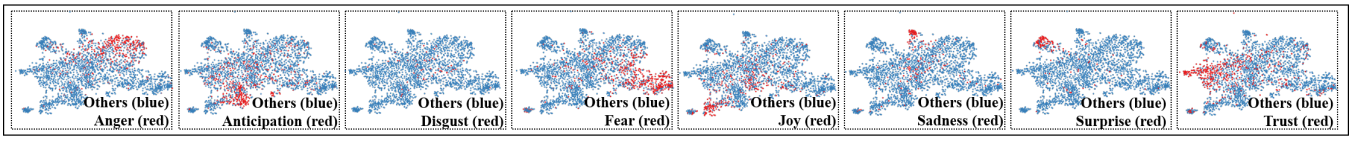
To evaluate the effectiveness of emotion-independent prior generation in our proposed adversarial disentangler (denoted as DR), we also prepare two variants: one simply removes the emotion-independent priors (denoted as $\text{DR}_{w/o}$); the other is with replacement of random noises (i.e., denoted as DR_{rnd}). Thus, we construct different variants by connecting different transformer based models with adversarial disentangler of different conditions. The DNN classifier used in Dr.Emotion is applied for all methods.

The experimental results are shown in Table 3, from which we can see that: (1) Directly using post embeddings obtained from different post-trained transformer based models (i.e., ID1, ID5, ID9, ID13) would achieve different performances for emotion analysis (i.e., RoBERTa, BERT and XLNet are comparable but much better than XLM). (2) Applying adversarial disentangler even with different variants improves emotion classification performance compared with their corresponding transformer based models without such mechanism (i.e., ID2-4 vs. ID1, ID6-8 vs. ID5, ID10-12 vs. ID9, ID14-16 vs. ID13). (3) The adversarial disentangler integrating the emotion-independent priors performs better than random condition denoted as DR_{rnd} and without condition denoted as $\text{DR}_{w/o}$ (i.e., ID4 vs. ID2-3, ID8 vs. ID6-7, ID12 vs. ID10-11, ID16 vs. ID14-15). (4) Dr.Emotion outperforms all others by achieving an impressive Micro-AUC of 0.9154, Macro-AUC of 0.9048, Micro-F1 of 0.7442 and Macro-F1 of 0.7311. The success of DR.Emotion may lie in: (1) it delicately post-trains the advantaged transformer based model (i.e., RoBERTa) to obtain initial post embeddings; (2) the proposed adversarial disentangler with incorporation of the generated emotion-independent priors is capable of separating implicitly expressed emotions from the content in latent space for emotion analysis, which can be further demonstrated by the following example.

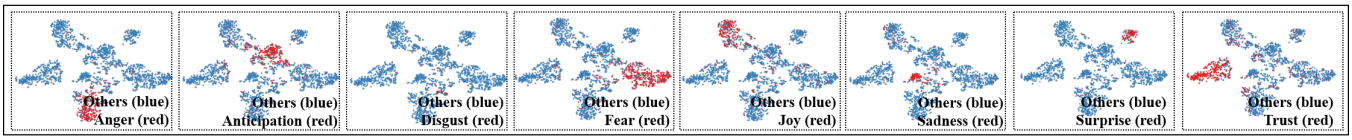
Tweet: "*Many nurse anesthetists in Pennsylvania have been laid off even though they are particularly critical to the coronavirus response because they can help intubate patients and manage them on ventilators. Huh? What's going on?*" This social media post is classified as *anticipation* by RoBERTa w/o disentanglement; while integrating the generated emotion-independent prior c=0.7046, it's classified as *anger* by Dr.Emotion.

Table 2: Evaluation of Dr.Emotion in emotion analysis of COVID-19 related social media posts (tweets).

ID	Model	AUC		F1		ID	Model	AUC		F1	
		Micro	Macro	Micro	Macro			Micro	Macro	Micro	Macro
1	RoBERTa	0.9031	0.8854	0.6652	0.6451	9	XLNet	0.8908	0.8864	0.6753	0.6544
2	RoBERTa+DR$_{w/o}$	0.9083	0.8922	0.6859	0.6944	10	XLNet+DR$_{w/o}$	0.8930	0.8821	0.6833	0.6605
3	RoBERTa+DR$_{rnd}$	0.9004	0.8863	0.6632	0.6359	11	XLNet+DR$_{rnd}$	0.8962	0.8891	0.6631	0.6439
4	**RoBERTa+DR (Dr.Emotion)**	**0.9154**	**0.9048**	**0.7442**	**0.7311**	12	XLNet+DR	0.9041	0.8913	0.6934	0.6787
5	BERT	0.8969	0.8763	0.6721	0.6452	13	XLM	0.8796	0.8607	0.5996	0.5703
6	BERT+DR$_{w/o}$	0.8993	0.8744	0.6874	0.6612	14	XLM+DR$_{w/o}$	0.8847	0.8669	0.6178	0.5831
7	BERT+DR$_{rnd}$	0.8987	0.8784	0.6674	0.6369	15	XLM+DR$_{rnd}$	0.8773	0.8521	0.5874	0.5553
8	BERT+DR	0.9058	0.8751	0.7142	0.6883	16	XLM+DR	0.8903	0.8689	0.6264	0.5933



(a) Results w/o Disentanglement (i.e., RoBERTa)



(b) Results with Disentanglement (i.e., Dr.Emotion)

Figure 5: Visualization of representations with and w/o disentanglement.

For a more intuitive comparison, we further visualize post representations without (i.e., directly obtained by post-trained RoBERTa) and with disentanglement (i.e., learned by Dr.Emotion), by projecting them into a two-dimension space through t-SNE algorithm [36]. We use a one-vs-rest manner (e.g., anger vs. non-anger, fear vs. non-fear, etc.) to illustrate the results. From the plots in Figure 5, we can see that DR.Emotion can successfully distinguish different types of emotions with denser clusters and more explicit boundaries.

3.3 Comparisons with Baseline Methods

In this section, we compare Dr.Emotion with various state-of-the-art methods for emotion analysis of social media posts, including traditional lexicon policy, n-gram models, skip-gram model as well as generative adversarial network (GAN) and variational autoencoder (VAE) based frameworks for disentangled representation learning in NLP. We here apply a three-layer MLP used in Dr.Emotion as the classification model for all baselines.

- **NRC lexicon**: The National Research Council Canada (NRC) emotion lexicon [37] is a dictionary mapping a word into a pair of emotion and intensity score. Given a post, using the lexicon policy, we add up the emotion scores of the words in the post to obtain its emotion category.
- **N-grams**: A post will be represented by 1, 2, 3-gram feature vectors fed to the classification model for emotion analysis.
- **Skip-gram model**: We exploit the skip-gram model GloVe [39] (i.e., GloVe.twitter.27B.200d) for comparison, which provides a pre-trained model that maps each word into a 200-dimension embedding. Given a post, we generate its feature vector by average pooling on all embeddings available in this post.
- **GAN-based framework** (denoted as GAN-based DR): In [23], an adversarial framework is proposed for text generation, where generator and discriminator compete with each other to retain the content while reducing stylistic properties. When applying such mechanism in our case, we replace the polar indicators by the generated emotion-independent priors.
- **VAE-based framework** (denoted as VAE-based DR): VAE-based models [2, 9, 25] have been studied for disentangled representation learning in text generation. Given a post embedding, we adapt a general VAE by feeding concatenation of sampled latent variable and the generated emotion-independent prior into the decoder; after optimization, the learned disentangled representation can be extracted before the concatenation operation.

The comparison results are shown in Table 3, from which we observe that: (1) compared with lexicon policy and skip-gram model, n-gram (i.e., unigram) representations perform better in emotion analysis of COVID-19 related tweets in our application (note that compared with AUC scores, the F1 scores have significant drops because emotion distributions in real-world social media data are quite unbalanced as shown in Table 1); (2) GAN-based and VAE-based frameworks (i.e., GAN-based DR and VAE-based DR) achieve better results than other models without disentangled representation learning mechanism; (3) DR.Emotion outperforms all baselines and significantly improves F1 scores.

Table 3: Comparisons of Dr.Emotion with baseline methods.

Model	AUC		F1	
	Micro	Macro	Micro	Macro
NRC Lexicon	0.7152	0.6948	0.4771	0.2855
Unigram	0.7855	0.8237	0.4927	0.4612
Bigram	0.7765	0.7231	0.3847	0.3584
Trigram	0.7495	0.6813	0.3228	0.2855
GloVe	0.7574	0.7089	0.3185	0.2629
GAN-based DR	0.9032	0.8844	0.6843	0.6810
VAE-based DR	0.9078	0.8741	0.7130	0.6774
Dr.Emotion	**0.9154**	**0.9048**	**0.7442**	**0.7331**

3.4 Stability, Sensitivity and Scalability

In this section, we fully evaluate the stability, parameter sensitivity and scalability of Dr.Emotion. We first examine the **stability** of Dr.Emotion by analyzing the training losses of generator and discriminator and convergence of the adversarial disentanlger. As shown in Figure 6.(a), all losses reach equilibrium after 15 epochs, which demonstrates the training stability of DR.Emotion; we also evaluate the F1 scores in terms of different training epochs and the results shown in Figure 6.(b) further proves that DR.Emotion achieves stable performance. For **parameter sensitivity** evaluation, we analyze how different choices of disentangled representation dimension (i.e. we vary the dimensions from 32 to 512) affect the performance of DR.Emotion. The results shown in Figure 6.(c) demonstrate that DR.Emotion is not strictly sensitive to the parameters and is able to reach optimal performance under a cost-effective parameter choice (i.e., when dimension is set to 128). For **scalability** evaluation, we investigate the training time of the adversarial disentangler of Dr.Emotion with different sizes of training data (i.e., from 5,000 to 27,999). The plots in Figure 6.(d) show that the disentangled representation learning is linear to the number of training data samples in each epoch.

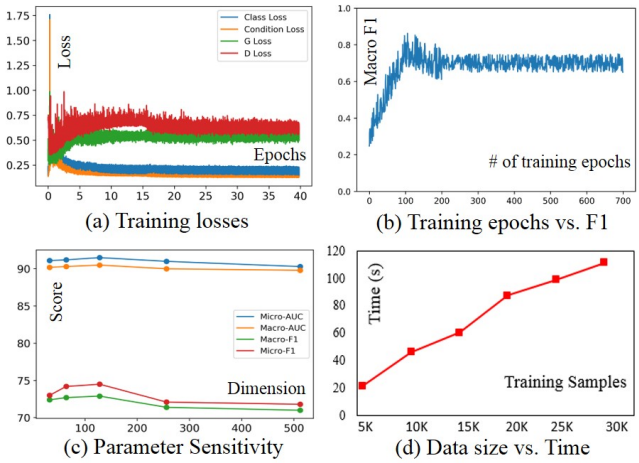


Figure 6: Stability, parameter sensitivity and scalability.

4 DEEP INVESTIGATION IN THE WILD

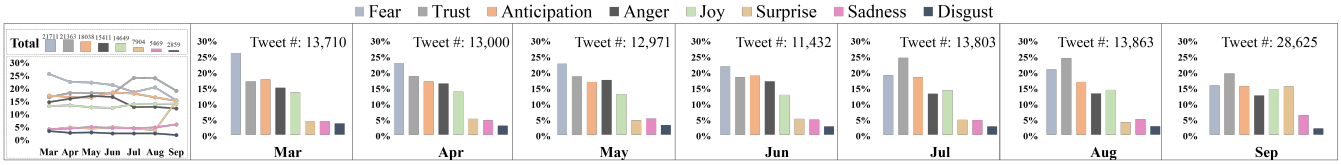
As discussed in Section 1, the prolonged pandemic has become a paradigm shifting phenomenon and exposed vulnerabilities impacting community resilience, including the inability to effectively and efficiently address the social, economic and behavioral issues. In response to the spread of COVID-19, many social activities have moved online; therefore, as the pandemic progresses, the information and emotions extracted from social media posts may play an important role to gain in-depth insights into public feelings for community resilience improvement. In this section, based on our large-scale data collected from social media (i.e., 142,434 COVID-19 related tweets posted by users in the U.S. through Mar 1, 2020 to Sep 30, 2020), by applying our proposed and developed *Dr.Emotion*, we perform the emotion analysis and further investigate public perceptions towards COVID-19 from both temporal and geographical perspectives to assist with community resilience improvement.

4.1 Temporal Analysis Exploiting Dr.Emotion

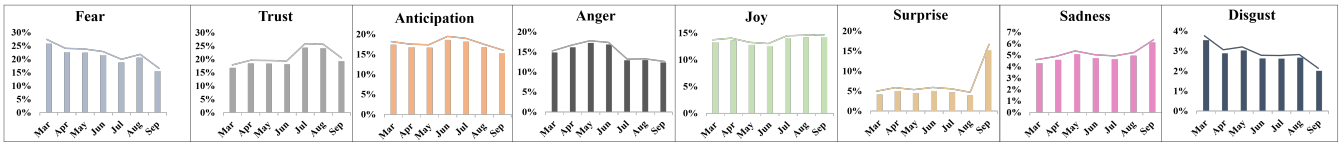
For temporal analysis, we use the labeled dataset (i.e., 27,999 tweets with manually labeled emotion classes) to train Dr.Emotion and then apply it to classify emotions of large-scale unlabeled tweets (i.e., 107,434 COVID-19 related tweets posted by users in the U.S. through Mar 1, 2020 to Sep 30, 2020). Based on the emotion classification results by Dr.Emotion, Figure 7.(a) shows the overall distributions of emotions and distributions in each month while Figure 7.(b) illustrates their changes over time; and Table 4 shows the top bigrams of tweets in each month and overall top bigrams. From Figure 7 and Table 4, we observe that: (1) Overall, the prolonged pandemic has caused dominant fear and volatile emotions in combination with anticipation and trust. (2) Since the lockdown orders initiated in Mar, the feelings of fear and anger combing anticipation and trust have been dominant till Jun. The reasons behind this could be perceived by the most discussed topics (e.g., *"epidemic, hard"*, *"getting infected"*, *"inconsistent income"*, *"suffering crisis"*, *"city lockdown"*, *"second wave"* and *"million patient"*), which indicate that the unseen pandemic not only physically endangers people, but also financially jeopardizes their daily life due to the lockdown and unemployment caused by COVID-19. (3) From Jul to Sep, fear mixed with anticipation and trust have been dominant feelings while anger decreased as people have been adapting themselves to the "new normal" and gaining more knowledge on COVID-19 (shown by topics such as *"making money"* and *"reopening school"*, particularly with surprise of *"rapid diagnostic"* in Sep). In summary, the dynamics between fear and anger mixed with anticipation and trust indicate that while people might have been worried less about the pandemic itself (e.g., *"epidemic ignorance"*), effects brought by COVID-19 such as substantial confirmed cases, fatalities, unemployment and quarantine fatigue drove the volatile emotions; nevertheless, starting from Jul, the feelings of trust have skyrocketed: from intensely discussed topics such as *"making money"*, *"wore mask"*, *"vaccine trail"*, and *"hired workers"*, it is noticed that our community has been trying to remigrate to "new normal" while fighting against the pandemic.

4.2 Geographical Analysis with Dr.Emotion

For geographical analysis, we select all social media posts with a U.S. location that can be pinpointed to a state during Sep 2020,



(a) Overall emotion distributions and distributions in each month through Mar 2020 to Sep 2020.



(b) Emotion changes over time through Mar 2020 to Sep 2020.

Figure 7: Temporal analysis of public emotions in social media (i.e., Twitter) by exploiting Dr.Emotion.

Table 4: Top bigrams extracted from large-scale social media posts (i.e., tweets).

Month	March	April	Month	May	June
bigram	(epidemic, hard) (getting, infected) (hit, stock) (lockdown, ban) (inconsistent, income)	(stayathome, order) (wash, hand) (industry, battered) (suffering, crisis) (city, lockdown)	bigram	(symptom, sign) (fda, emergency) (nurse, together) (epidemic, ignorance) (social, distancing)	(second, wave) (health, department) (outdoor, spread) (police, violence) (million, patient)
Month	July	August	Month	September	Total
bigram	(making, money) (continue, climb) (staysafe, mask) (hope, treatment) (covid, party)	(wore, mask) (help, flattenthecurve) (vaccine, trial) (reopening, school) (pay, rent)	bigram	(herd, immunity) (stimulus, bill) (rapid, diagnostic) (hired, workers) (wait, vaccine)	(highest, risk) (social, distancing) (health, crisis) (wore, mask) (infected, recover)

which result in a total number of 28,623 tweets and overall with leading emotions of trust, fear and anticipation.

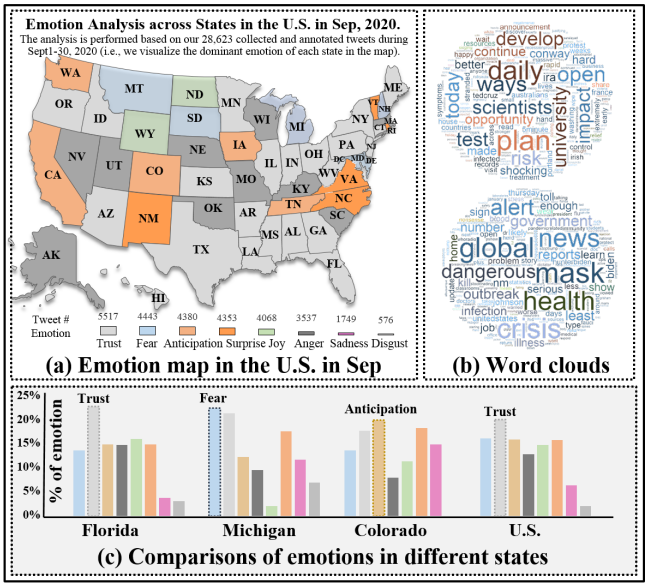


Figure 8: Primary emotion of each state within the U.S.

The visualization of dominant emotion in each state shown in Figure 8.(a) illustrates that the states within the U.S. are dominated by trust in combination with volatile emotions. Figure 8.(b) shows the word clouds generated for anticipation (top) and fear (bottom) respectively. From the cloud of anticipation, we can see keywords *"scientists", "impact", "plan", "daily", "opportunity"* mostly appear, which might infer that the public anticipation during Sep mainly lies on virus precaution, scientific solution and long-term objectives. As shown by intensely discussed topics in Table 4, people at the beginning of this pandemic mainly focused on imminent ripples directly brought by the pandemic; however, after almost seven months of adaptation, what people really anticipate may have shifted from solving emergencies to chronically combating and soothing the genuine pandemic itself. From the cloud of fear, we discover that topics containing *"mask", "health", "crisis", "global", "dangerous", "alerts"* and *"news"* are the ones that may cause feelings of fear. The observation may indicate that i) the source of people's fear has not changed since the beginning of COVID-19; ii) ripples brought by the pandemic such as medical supply shortage, unemployment, and health precaution still significantly perturb people's mentality. However, even though the total number of fear emotion is still relatively high, only few states are dominated by fear during Sep, which may imply that people might have been positively adapting to this global crisis. In Figure 8.(c), we select three states for deeper analysis: Florida, Michigan and Colorado where

the severity of novel disease is severe, moderate and mild. For each of them, we query the keywords related to the dominated emotion, from which we discover that: i) in Florida which is one of the most severe states, people mainly feel trust with most discussed topics of "vaccine", "safety", "facility" and "student"; ii) in Michigan with moderate situation, people mainly feel fear with topics of "suffering", "economic", "doctors" and "mask"; iii) in Colorado where the crisis is mild, people mostly feel anticipation with intensely discussed topics of "proposal", "community", "strength" and "battling". By acknowledging what a subgroup of people truly anticipate or fear, actions could be implemented to mitigate negative ripples by policy makers, public health experts, business owners, organizations (e.g., schools), and any individual of interest.

5 LIMITATIONS AND DISCUSSION

Although our proposed *DR.Emotion* could obtain desirable results, it is also subject to certain limitations: data quality and analysis bias. For data quality, unlike conventional literature texts, social media data, especially for Twitter data, are highly noisy. There exist misspellings and punctuation errors in the data and online users incline to use abbreviations and non-dictionary slang. Moreover, the quality of data can also be affected by bot activity, repeat posts and spam posts. In our work, although we have exploited preprocessing techniques to deal with the large-scale data collected from Twitter, further data cleaning and preprocessing may help further improve the data quality. For analysis bias, as using entire Twitter posts for analysis is infeasible, we perform our investigation based on our established dataset; although it's sampled dataset, the large number of posts (i.e., 142,434 tweets from Mar to Sep) could be representative to some extent. Meanwhile, our geographical analysis is based on the Twitter service to acquire the geolocation information of user posts (note that all users are anonymized in our work using hash values of usernames). However, such information is often missing, unstructured, or nongeographical and thus only a portion of user geolocation can be precisely obtained. This fact leads to a scarce data collection for some states; thus, the distribution of our sampled data may diverge from that of the true data, which would yield biased analysis results. Despite these limitations, as initial efforts, our work in this paper may provide in-depth insights and customized guidance that can help public health experts, social workers and policy makers in decision-making and also enable a conceptual framework for the development of resilient community engagement strategies in responses to a variety of crises created by COVID-19 and future natural or health-related disasters.

6 RELATED WORK

Emotion/Sentiment Analysis. There have been many research efforts on emotion/sentiment analysis of texts. Traditional computational methods for identifying emotions/sentiments include classic machine learning algorithms (e.g., naive Bayes [18], SVM [50], logistic regression [35]) based on handcrafted features (e.g., lexicons) or feature engineering techniques (e.g., *n*-grams). In recent years, deep learning models such as CNNs [28, 54], RNNs [34, 47], Transformer [46] have outperformed the traditional machine learning models for this task. Thanks to the success of transformer based models such as BERT [12], RoBERTa [32], XLNet [52] and XLM [30], the methods proposed to post-train these models [13, 27, 51, 53] have achieved state-of-the-art performance in various NLP tasks. To access public feelings during the pandemic, there have been research studies of emotion/sentiment analysis in terms of COVID-19 [14, 15, 29, 40, 41]. For example, [41] analyzes geographical sentiment distributions using traditional machine learning algorithms; [15] specifically investigates college students' responses to the pandemic with RNNs. Different from these works, in this paper, based on our established large-scale COVID-19 related social media dataset with credible ground-truth, we focus on disentangled representation learning in social media posts for emotion analysis and further investigate public perceptions towards COVID-19 from both temporal and geographical perspectives to assist with community resilience improvement.

Disentangled Representation Learning in NLP. Disentangled representation learning has shown its success in computer vision domain [8, 17, 20, 43]. It has also inspired researches in the NLP field [2, 3, 7, 9, 11, 23, 25, 31], most of which focus on text generation. To put this into perspective, [2, 3, 7] propose to disentangle syntactic and semantic representations from a given sentence; in particular, [7] applies a deep generative model consisting of von Mises Fisher and Gaussian priors on the semantic and syntactic latent variables and a deep bag-of-words decoder that conditions on these latent variables. [11, 23, 25, 31] learn the latent space that is disentangled to retain text content while reducing stylistic properties. As users may implicitly express their emotions in social media posts which could be highly entangled with the content, to address this challenge for emotion analysis, different from the existing works, we propose an adversarial disentangler by integrating emotion-independent priors of the posts generated by a post-trained transformer-based model to separate and disentangle the implicitly encoded emotions from the content in latent space *at the first attempt*.

7 CONCLUSION

To combat the prolonged pandemic that has exposed vulnerabilities impacting community resilience, in this paper, based on our established large-scale COVID-19 related social media data, we propose and develop an integrated framework Dr.Emotion to learn disentangled representations of social media posts (i.e., tweets) for emotion analysis and thus to gain insights into public feelings of COVID-19. In Dr.Emotion, for each given social media post, we first post-train a transformer-based model to obtain the initial post embedding; and then by integrating its generated emotion-independent prior, an adversarial disentangler is proposed to separate and disentangle the embedded emotions from the content in latent space for emotion classification. Extensive experimental studies and promising results demonstrate the performance of Dr.Emotion in tweet emotion analysis by comparisons with state-of-the-art baselines. By exploiting our developed Dr.Emotion, we further perform emotion classification based on 107,434 COVID-19 related tweets posted by users in the U.S. through Mar to Sep and provide in-depth investigation from both temporal and geographical perspectives, based on which additional work can be conducted to extract and transform the constructive ideas, experiences and support into actionable information to improve community resilience in responses to a variety of crises created by COVID-19 and well beyond.

REFERENCES

[1] Anonymous. 2020. *Sample dataset and open-source codes of Dr.Emotion.* https://github.com/www2021DrEmotion/www2021DrEmotion.

[2] Vikash Balasubramanian, Ivan Kobyzev, Hareesh Bahuleyan, Ilya Shapiro, and Olga Vechtomova. 2020. Polarized-VAE: Proximity Based Disentangled Representation Learning for Text Generation. *arXiv preprint arXiv:2004.10809* (2020).

[3] Yu Bao, Hao Zhou, Shujian Huang, Lei Li, Lili Mou, Olga Vechtomova, Xinyu Dai, and Jiajun Chen. 2019. Generating Sentences from Disentangled Syntactic and Semantic Spaces. In *ACL*. 6008–6019.

[4] Alexander W. Bartik, Marianne Bertrand, Zoë B. Cullen, Edward L. Glaeser, Michael Luca, and Christopher T. Stanton. 2020. How are small businesses adjusting to covid-19? early evidence from a survey. *National Bureau of Economic Research* (2020).

[5] BEA. 2020. *Gross Domestic Product (Third Estimate), Corporate Profits (Revised), and GDP by Industry, Second Quarter 2020.* https://www.bea.gov/news/2020/gross-domestic-product-third-estimate-corporate-profits-revised-and-gdp-industry-annual.

[6] BLS. 2020. *The employment situation - May 2020.* https://www.bls.gov/news.release/pdf/empsit.pdf.

[7] Mingda Chen, Qingming Tang, Sam Wiseman, and Kevin Gimpel. 2019. A Multi-Task Approach for Disentangling Syntax and Semantics in Sentence Representations. In *NAACL*. 2453–2464.

[8] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *NIPS*. 2172–2180.

[9] Xingyi Cheng, Weidi Xu, Taifeng Wang, Wei Chu, Weipeng Huang, Kunlong Chen, and Junfeng Hu. 2019. Variational Semi-Supervised Aspect-Term Sentiment Analysis via Transformer. In *CoNLL*. 961–969.

[10] E. H. Coe and K. Enomoto. 2020. *Returning to resilience: The impact of COVID-19 on mental health and substance use.* https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/returning-to-resilience-the-impact-of-covid-19-on-behavioral-health.

[11] Ning Dai, Jianze Liang, Xipeng Qiu, and Xuan-Jing Huang. 2019. Style Transformer: Unpaired Text Style Transfer without Disentangled Latent Representation. In *ACL*. 5997–6007.

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

[13] Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, and Jianxin Liao. 2020. Adversarial and Domain-Aware BERT for Cross-Domain Sentiment Analysis. In *ACL*. 4019–4028.

[14] Akash Dutt Dubey. 2020. Decoding the Twitter Sentiments towards the Leadership in the times of COVID-19: A Case of USA and India. *SSRN:3588623* (2020).

[15] Viet Duong, Phu Pham, Tongyu Yang, Yu Wang, and Jiebo Luo. 2020. The ivory tower lost: How college students respond differently than the general public to the covid-19 pandemic. *arXiv preprint arXiv:2004.09968* (2020).

[16] Emilien Dupont. 2018. Learning disentangled joint continuous and discrete representations. In *NIPS*. 710–720.

[17] Emilien Dupont. 2018. Learning disentangled joint continuous and discrete representations. In *NIPS*. 710–720.

[18] Pablo Gamallo and Marcos Garcia. 2014. Citius: A NaiveBayes Strategy for Sentiment Analysis on English Tweets. In *SemEval*. Citeseer.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *NIPS*. 2672–2680.

[20] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2016. beta-vae: Learning basic visual concepts with a constrained variational framework. In *ICLR*.

[21] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2017. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. *ICLR* 2, 5 (2017), 6.

[22] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. *A field guide to dynamical recurrent neural networks* (2001).

[23] Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. 2017. Toward controlled generation of text. In *ICML*. 1587–1596.

[24] JHU. 2020. *Coronavirus COVID-19 Global Cases.* https://coronavirus.jhu.edu/map.html.

[25] Vineet John, Lili Mou, Hareesh Bahuleyan, and Olga Vechtomova. 2019. Disentangled Representation Learning for Non-Parallel Text Style Transfer. In *ACL*. 424–434.

[26] Heather J. Kagan. 2020. *Opioid overdoses on the rise during COVID-19 pandemic, despite telemedicine care.* https://abcnews.go.com/Health/opioid-overdoses-rise-covid-19-pandemic-telemedicine-care/story?id=72442735.

[27] Pei Ke, Haozhe Ji, Siyang Liu, Xiaoyan Zhu, and Minlie Huang. 2020. SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge. *arXiv preprint arXiv:1911.02493* (2020).

[28] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).

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

[30] Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291* (2019).

[31] Maria Larsson, Amanda Nilsson, and Mikael Kågebäck. 2017. Disentangled representations for manipulation of sentiment in text. *arXiv preprint arXiv:1712.10066* (2017).

[32] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[33] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. 2019. Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations. In *ICML*. 4114–4124.

[34] Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM.. In *AAAI*. 5876–5883.

[35] Andrew Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *ACL: HLT*. 142–150.

[36] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, Nov (2008), 2579–2605.

[37] Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Council, Canada* 2 (2013).

[38] NY-Governor. May 20, 2020. *Following Spike in Domestic Violence During COVID-19 Pandemic, Secretary to the Governor Melissa Derosa & NYS Council on Women & Girls Launch Task Force to Find Innovative Solutions to Crisis.* https://www.governor.ny.gov/news/following-spike-domestic-violence-during-covid-19-pandemic-secretary-governor-melissa-derosa.

[39] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *EMNLP*. 1532–1543.

[40] Atika Qazi, Javaria Qazi, Khulla Naseer, Muhammad Zeeshan, Glenn Hardaker, Jaafar Zubairu Maitama, and Khalid Haruna. 2020. Analyzing situational awareness through public opinion to predict adoption of social distancing amid pandemic COVID-19. *Journal of Medical Virology* (2020).

[41] Jim Samuel, GG Ali, Md Rahman, Ek Esawi, Yana Samuel, et al. 2020. Covid-19 public sentiment insights and machine learning for tweets classification. *Information* 11, 6 (2020), 314.

[42] Spotcrime. June, 2020. *Daily Crime Blotter in Chicago.* https://spotcrime.com/il/chicago/daily.

[43] Luan Tran, Xi Yin, and Xiaoming Liu. 2017. Disentangled representation learning gan for pose-invariant face recognition. In *CVPR*. 1415–1424.

[44] Twitter. 2020. *Twitter API.* https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.

[45] Vincent Van Asch. 2013. Macro-and micro-averaged evaluation measures [[basic draft]]. *Belgium: CLiPS* 49 (2013).

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*. 5998–6008.

[47] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based LSTM for aspect-level sentiment classification. In *EMNLP*. 606–615.

[48] WHO. 2020. *Coronavirus disease (COVID-19).* https://www.who.int/.

[49] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *EMNLP*. 347–354.

[50] Rui Xia, Chengqing Zong, and Shoushan Li. 2011. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences* 181, 6 (2011), 1138–1152.

[51] Hu Xu, Bing Liu, Lei Shu, and Philip S Yu. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232* (2019).

[52] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *NIPS*. 5753–5763.

[53] Da Yin, Tao Meng, and Kai-Wei Chang. 2020. SentiBERT: A Transferable Transformer-Based Architecture for Compositional Sentiment Semantics. *arXiv preprint arXiv:2005.04114* (2020).

[54] Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. In *NIPS*. 649–657.