Emotion-Trigger Summarization: A Computational Study on COVID-19 Social Media

Hongli Zhan

Abstract

Crises such as the COVID-19 pandemic continuously threaten our world and emotionally affect millions of people worldwide in distinct ways. Understanding the triggers leading to people's perceived emotions is of crucial importance. In this paper, we propose a new task *Emotion-Trigger Summarization*, whose objective is to abstractively summarize triggers with respect to automatically detected emotions in a document. By introducing abstractive summarization to emotion triggers, our task innovatively integrates emotion triggers that are otherwise multi-faceted and require the context information for interpretation. To address the new task, we introduce a benchmark dataset of ~1,900 Reddit posts in English from a Reddit forum r/COVID19_support annotated for emotions and triggers in posts. Our dataset makes up for the vacancy of an emotion trigger dataset in English. Based on our dataset, we propose a joint neural network model combining both emotion detection and emotion-based trigger summarization. We release our dataset and codes at https://github.com/honglizhan/Emotion-Trigger (currently private).

Reddit Post

1: My sibling is 19 and she constantly goes places with her friends and to there houses and its honestly stressing me out.
2: Our grandfather lives with us and he has dementia along with other health issues and my mom has diabetes and heart problems and I have autoimmune diseases & chronic health issues.
3: She also has asthma.
4: Its stressing me out because despite this she seems to not care about how badly it would affect all of us if we were to get the virus.
5: And sadly I feel like its not much I can do she literally doesn't respect my mom and though I'm older she doesn't respect me either.
6: Its so frustrating.

Emotions and Abstractive Summaries of Triggers

Emotion: anger
Abstractive Summary of Trigger: My sister having absolutely no regard for any of our family's health coupled with the fact that I can't do anything about it is so aggravating to me.

Emotion: fear
Abstractive Summary of Trigger: My sibling, who, in spite of our family's myriad of issues that all make us high-risk people, continuously goes out and about, which makes her likely to get infected. I am scared for all of us right now.

Figure 1: An example from our dataset, with colors indicating emotions. Note that due to the space limit, we only show here the annotations from one annotator. In the benchmark dataset each example is annotated by two annotators.

1 Introduction

Emotion analysis is the area of research which encompasses emotion detection, polarity classification, and cause detection (Khunteta and Singh, 2021). Early research in emotion analysis focused on emotion detection (Alm et al., 2005; Xu et al., 2012), whilst recent work has gone beyond emotion classification to uncover the causes behind emotions. Emotion Cause Extraction is a thriving task in the past decade that aims to extract the events triggering a particular emotion (Khunteta and Singh, 2021). Lee et al. (2010) first introduced the task of Emotion Cause Extraction on word-level causes, and Chen et al. (2010) expanded the task to clause-level cause detection. Later, Gui et al. (2016) redefined the Emotion Cause Extraction task as a clause-level binary classification problem. Xia and Ding (2019) captured the shortcoming that emotions must be human-annotated before conducting automatic cause extraction and further proposed the task of Emotion-Cause Pair Extraction aiming to extract potential pairs of emotions and causes in a document.

However, we notice two shortcomings in the current task of Emotion Cause Extraction. First, there may be multiple causes associated with a single emotion in the document, as we showcase in Figure 1. The current task of Emotion Cause Extraction has yet to consider the intrinsic relations and the integration of multiple causes. Second, an extracted cause from the document may be hard to understand without the context information. For example, in Figure 1, the sentence "Despite this she seems to not care about how badly it would affect all of us if we were to get the virus" is the cause for "anger". However, due to the use of pronouns, it is impossible for the audience to acquire the entire perspective from the sentence.

In this paper, we propose a new task *Emotion-Trigger Summarization*, which aims to abstractively summarize triggers with respect to automatically detected emotions in a document. By introducing abstractive summarization to emotion triggers, we innovatively resolve the above-mentioned defects. In the present study, we refer to the events leading to a particular emotion in a document using the term *trigger* instead of *cause*. We argue this from a psychological standpoint: according to the APA Dictionary of Psychology[1], *trigger* is defined as "a stimulus that elicits a reaction", whereas *cause* is defined as "an event or state that brings about another". Thus, referring to the events leading to emotions as *trigger* is more literally accurate in this paper.

To establish the task of Emotion-Trigger Summarization, we present a novel emotion-trigger dataset in English from Reddit posts on COVID-19. The dataset is annotated with the Plutchik-8 emotions (Plutchik, 2001), where each emotion is provided with an intensity label from the fine-grained Plutchik-24 labels. The trigger(s) leading to each emotion are given in both extractive and abstractive forms, in which the extractive summaries are highlights of emotion triggers from the Reddit post and the abstractive summaries are highly-condensed, integrated sentences of emotion triggers. We showcase examples of our dataset in Appendix §C.

Our dataset makes up for the vacancy that the corpora discovered in the literature review of uncovering triggers behind emotions are exclusively in Chinese (Khunteta and Singh, 2021; Drury et al., 2022) and relatively small-in-size (Wang et al., 2020). Besides, existing work has focused on emotions that are explicitly expressed by keywords, and prior research has yet to take into account implicit emotions that are implied and require reasoning (Khunteta and Singh, 2021). As we demonstrate in §5, our dataset responds to the need of implicit emotions in the field.

Drawing from studies in aspect-based as well as query-based summarization, we propose an emotion-based trigger summarization approach to address the task of Emotion-Trigger Summarization based on our dataset. Unlike previous works, we perform trigger summarization with regard to automatically detected emotions in the document.

[1]https://dictionary.apa.org/

We explore the detection of perceived emotions and the summarization of triggers behind the emotions in social media on COVID-19 to answer two research questions. First, what are the emotions expressed through social media on subjects related to COVID-19? How do the emotions change over time? What are the triggers leading to the emotions? Second, how well can we benchmark our dataset on modern large-scale pre-trained NLP models for the tasks of emotion detection and emotion-based trigger summarization?

We summarize our contributions as follows:

- We propose a new task *Emotion-Trigger Summarization*, which aims to abstractively summarize triggers with respect to automatically detected emotions in a document. The new task successfully takes into account triggers that are multi-faceted and require the context information for interpretation.

- We introduce an emotion-trigger dataset including an expert-annotated benchmark that makes up for the vacancy of an emotion trigger dataset in English. Unlike previous literature, our dataset also considers implicit emotions that require implication.

- We propose an emotion-based trigger summarization approach based on our dataset to address the task of Emotion-Trigger Summarization. Conditioned on automatically detected emotions in a document, triggers eliciting the emotions are abstractively summarized.

This paper is organized in the following way. §3 discusses the construction of our dataset. We present the inter-annotator agreement statistics in §4, and in §5 we provide a detailed qualitative analysis of our dataset. In §6 we benchmark several state-of-the-art emotion detection and automatic text summarization models on our dataset.

2 Background and Related Work

2.1 Emotion Cause Extraction

Emotion Cause Extraction is a task that aims to extract the events triggering a particular emotion (Khunteta and Singh, 2021). Emotion Cause Extraction was first introduced by Lee et al. (2010), where they defined the task as extracting word-level causes to the given emotion in text. Chen et al. (2010) suggested that a clause may be the most appropriate unit for cause detection and expanded the task from word-level to clause-level cause detection. Later, Gui et al. (2016) redefined the Emotion Cause Extraction task as a clause-level binary classification problem.

However, in the above task settings, emotions must be human-annotated before conducting automatic cause extraction. Xia and Ding (2019) captured this shortcoming, and further proposed the task of Emotion-Cause Pair Extraction, which is defined as extracting potential pairs of emotions and causes in a document.

Unlike the above works, in this paper we propose a new task *Emotion-Trigger Summarization*, which aims to abstractively summarize triggers with respect to automatically detected emotions in a document. In the present study, we refer to the events leading to a particular emotion in a document using the term *trigger* instead of *cause*. To our knowledge, this is the first study to perform abstractive summarization on emotion triggers. Our new task is similar to Emotion-Cause Pair Extraction (Xia and Ding, 2019) in that we do not require emotions annotated before trigger summarization, thus making the task more applicable in reality. By introducing abstractive summarization to emotion triggers, our work innovatively integrates triggers that are otherwise spread across the document. Besides, given that one usually needs the context to comprehensively understand an event, our new task also enhances the interpretability towards emotion triggers.

2.2 Existing Datasets

Several corpora have been constructed for the task of Emotion Cause Extraction. Lee et al. (2010) constructed a Chinese emotion corpus from 6,058 entries of Sinica data based on emotion keywords. Gao et al. (2015) built a Chinese corpus from 18,000 micro-blog posts. Gui et al. (2016) presented a dataset in Chinese that was extracted from SINA city news using emotion keywords. There were 15,687 instances in their corpus and each instance contained only one emotion marked by keywords and one or more annotated causes. This dataset serves as the benchmark for many following studies on Emotion Cause Extraction (Gui et al., 2017; Xia and Ding, 2019; Xiao et al., 2019; Fan et al., 2019; Xu et al., 2019; Hu et al., 2019).

Despite its popularity, Emotion Cause Extraction remains a challenging task due to the lack of datasets available (Wang et al., 2020). The corpora discovered in the literature review are exclusively in Chinese (Khunteta and Singh, 2021; Drury et al., 2022) and relatively small-in-size (Wang et al., 2020). Besides, existing work has focused on the extraction of causes behind explicit emotions that are expressed by keywords, while ignoring implicit emotions that are implied and require reasoning (Khunteta and Singh, 2021).

In the current study, to address our task of Emotion-Trigger Summarization, we introduce a novel dataset in English sourced from Reddit posts during COVID-19, where each post is human-annotated with emotions and triggers. We use the Plutchik-8 primary emotions as our emotion taxonomy, and we also provide intensity labels for each emotion using the fine-grained Plutchik-24 emotions (Plutchik, 2001). Unlike previous datasets on Emotion Cause Extraction which only consider explicit emotions, we demonstrate in our qualitative analysis of the dataset (§5) that the emotions in our dataset are considerably implicit.

2.3 Methodology

Previous literature have addressed Emotion Cause Extraction in various ways. Researchers have resorted to rule-based methods (Chen et al., 2010; Neviarouskaya and Aono, 2013; Gao et al., 2015; Yada et al., 2017) as well as traditional machine learning approaches (Russo et al., 2011; Gao et al., 2013; Li and Xu, 2014; Li et al., 2015; Ghazi et al., 2015; Gui et al., 2016; Xu et al., 2017). More recently, more studies have taken to deep neural networks to tackle Emotion Cause Extraction (Chen et al., 2018; Li et al., 2018; Xiao et al., 2019; Fan et al., 2019; Xia and Ding, 2019; Xu et al., 2019; Ding et al., 2020; Chen et al., 2020; Fan et al., 2020; Chen et al., 2022).

In the present study, we propose a joint deep neural model combining both emotion detection and automatic text summarization to address our new Emotion-Trigger Summarization task.

Emotion Detection. The task of emotion detection has been the focus in early works on Emotion Analysis, and researchers have extensively explored emotion classification in songs (Mihalcea and Strapparava, 2012), classic literature (Liu et al., 2019), online news (Strapparava and Mihalcea, 2007; Lei et al., 2014), health-related blog posts and online communities (Khanpour and Caragea, 2018; Sosea and Caragea, 2020), general social media domains (Wang et al., 2012; Abdul-Mageed and Ungar, 2017; Demszky et al., 2020), and natural disaster centric tweets (Desai et al., 2020).

Automatic Text Summarization. Automatic text summarization refers to the task of producing a compact summary that retains the essential information in the original text (Allahyari et al., 2017; Mridha et al., 2021). Automatic text summarization can be achieved through either extractive, abstractive, or hybrid means (El-Kassas et al., 2021). Extractive summarization identifies the most important sentences in the original text as the output summary, whilst abstractive summarization often requires language generation abilities to produce paraphrases of the original text (Allahyari et al., 2017; Liu and Lapata, 2019; El-Kassas et al., 2021).

Among the various types of automatic text summarization systems, query-based summarization and aspect-based summarization are two sub-tasks most similar in setting to the emotion-based trigger summarization approach we employ in this study. Query-based summarization aims to produce summaries from the document that are most relevant to the given search query (Zhong et al., 2021; El-Kassas et al., 2021). On the other hand, the objective of aspect-based summarization is to generate summaries of the document with respect to different aspects or perspectives (Tan et al., 2020; Hayashi et al., 2021).

Drawing from query-based and aspect-based summarization techniques, we propose the emotion-based trigger summarization approach to address the current task of Emotion-Trigger Summarization. Different from the above techniques, we perform trigger summarization with regard to automatically detected emotions in the document.

2.4 Reddit

Reddit is a popular online social forum that ranks the 7th most visited site in the United States[2]. It has 52 million daily active users who gather in one or more of over 2.5 million different communities called *subreddits* to connect with others on a specific interest or topic. Reddit does not impose short length limits on posts, and users can also comment on posts and engage in the conversations. Despite the intrinsic conversational nature of Reddit, few datasets in emotion analysis are sourced from Reddit. Among them, GoEmotions released by Demszky et al. (2020) is an emotion dataset from Reddit comments annotated with 27 emotions or Neutral.

Reddit has been used extensively for the study of mental well-being (Shen and Rudzicz, 2017; Pirina and Çöltekin, 2018; Turcan and McKeown, 2019; Jiang et al., 2020; Ashokkumar and Pennebaker, 2021; Seraj et al., 2021). Choudhury and De (2014) examined the mental health discourse on Reddit, and concluded that users share experiences around their illness challenges in their personal as well as professional lives and seek diagnosis, treatment, and emotional support for their conditions. In their research, Gjurković and Šnajder (2018) also came to the conclusion that Reddit is a gold mine for personality predictions.

3 Dataset Construction

In this section, we describe the collection and annotation processes of our dataset. We first discuss the source of our data in §3.1, then we introduce the annotation task on the Amazon Mechanical Turk in §3.2. Finally, we report our benchmark dataset in §3.3.

3.1 Selecting & Curating Reddit Posts

In this paper, we present a novel dataset from English Reddit posts that is manually annotated with emotions and triggers. We use the PSAW wrapper for Pushshift API[3] to gather Reddit posts in English from the subreddit r/COVID19_support[4], a Reddit forum created in February 2020. r/COVID19_support is chosen as the source of our data because of its rich personal narration: rather than COVID-19 news snippets, the subreddit is targeted for people seeking support during the pandemic. We sample posts before and after Omicron[5], a COVID-19 variant that emerged during December 2021.

Data Preprocessing. We gather posts 50-400 tokens long (punctuation excluded) by applying regular expressions. Web links are masked with an `[url]` token. The metadata are not provided to annotators. Details of the preprocessing procedure are provided in Appendix §A.

[2]https://websitebuilder.org/blog/reddit-statistics/

[3]https://psaw.readthedocs.io/en/latest/

[4]https://www.reddit.com/r/COVID19_support/

[5]https://www.cdc.gov/coronavirus/2019-ncov/variants/omicron-variant.html

3.2 Annotation Task

We crowdsource emotion-trigger annotations on the Amazon Mechanical Turk. We assign two annotators to each example.

Annotators. We recruit two different groups of annotators. The first group consists of turkers from the Amazon Mechanical Turk. The turkers all reside in the United States and have completed 500+ HITs with an acceptance rate $\geq 95\%$. For the second group of annotators, we recruit 2 undergraduate students from the Department of Linguistics at the University of Texas at Austin. We consider them as annotators with an expertise in linguistics for the present study.

We provide a survey on the annotators' personality in Appendix §B.1. Each group of annotators is responsible for the annotation of a disparate subset of the dataset. There are no mixed annotations in which a post is cross-annotated by two annotators from different groups. To ensure the quality of the dataset, both groups of annotators are trained in a *pre-annotation* process. Only the annotators who reach our expectations earn the qualification to participate in the annotation task. We also ask them to revise their work when needed during annotation. See Appendix §B.3 and §B.4 for more details.

Instructions. All annotators are given the same instructions. Specifically, annotators are asked to annotate perceived Plutchik-8 primary emotions (*anger, anticipation, joy, trust, fear, surprise, sadness, disgust*). Multiple selection is allowed, and we also provide a *none of the above* option in case no emotion is perceived. Once the annotators perceive an emotion, they are asked to select the intensity of the emotion from Plutchik-24 labels. There are 3 intensity levels for each Plutchik-8 emotion (e.g., the intensities for the emotion *anger* are: *annoyance, anger, rage*, as shown in Figure 2), and the annotators are allowed to choose only one intensity label per emotion.

We also ask the annotators to annotate the trigger(s) to their perceived emotions. The annotators are required to provide the extractive as well as abstractive summaries of the trigger(s) for each emotion they perceive in the post. The summaries should contain trigger(s) to the emotion rather than just reflecting the emotion itself. For the extractive summary of trigger, annotators need to provide the clause or sentence in the post that best describes the trigger of the emotion. Since it is possible that

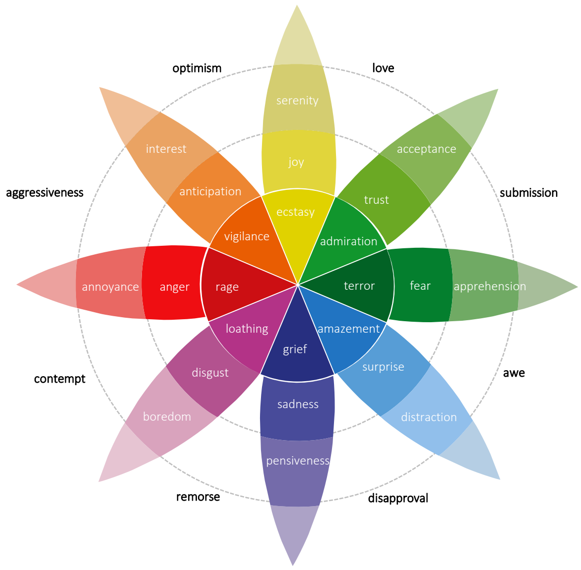

Figure 2: Plutchik's Wheel of Emotions. The 8 primary emotions lie in the middle circle of the wheel. Colors indicate the intensity of emotions: the darker the shade, the more intense the emotion.

there are multiple triggers contributing to the same emotion in a document, we ask annotators to separate the sentences if they are drawn from different parts of the post. For the abstractive summary of trigger, annotators are asked to summarize in their own words the trigger(s) of the emotion from the perspective of the poster. The abstractive summary is a free-form annotation that should integrate and reflect the trigger(s) of the perceived emotion in a highly condensed way. We provide detailed examples to help annotators navigate our trigger summarization task. We provide the detailed instructions to our annotation task in Appendix §B.2.

3.3 Benchmark Dataset

We construct a benchmark dataset to facilitate the modelling for the task of Emotion-Trigger Summarization. We aggregate all the emotion labels perceived by annotators in each post, and we also keep all the annotated trigger summaries for future multi-reference model evaluations.

Next, we create the training, validation, and test sets. As we have introduced above, the two groups of annotators in this study each annotates a disparate subset of our dataset, and there are no cases where a post is cross-annotated by annotators from different groups. Therefore, we first gather all the data annotated by the linguistic experts and construct the test set accordingly. We further perform a 80/20 split on the rest of the data annotated by the turkers to create the training and validation sets. In the end, we have 1200 examples in the training set, 300 examples in the validation set, and 400 examples in the test set.

Through systematically separating the annotators in the test set from the annotators in the training data, our dataset to some extent eliminates the intrinsic bias introduced by the same annotators, thus making the test set more "unseen" to the models. Besides, a test set annotated by experts in linguistics also adds to the reliability of our benchmark dataset.

4 Inter-Annotator Agreement

In this section, we provide the inter-annotator agreement statistics of our dataset. We first report the emotion agreement statistics in §4.1, then we evaluate the trigger agreement for matching emotions in §4.2. Finally, we report results of human validation for a subset of our training data in §4.3. We also present a summary of the statistics in Table 1.

4.1 Agreement in Emotions

As shown in Table 1, 81.8% of the examples in our dataset have at least 1 emotions agreement between annotators, and 28.1% of the examples have at least 2 emotions agreement.

PEA Score. We first employ the Plutchik Emotion Agreement (PEA) metric introduced by Desai et al. (2020) to determine the inter-annotator agreement of emotions in our dataset. PEA measures agreement between Plutchik emotions based on their relative distance on the wheel. The PEA score ranges from 0 to 1, with a higher score indicating higher agreement[6].

We report here the average PEA score among annotators weighted by their numbers of annotations. Results show that the weighted average PEA score among workers in the entire dataset is 0.803, suggesting fairly strong inter-annotator agreement. More specifically, the PEA score of the training/-validation sets combined is 0.798, whereas the PEA score of the test set is 0.823. This reflects considerably higher quality in emotion annotations by the expert linguists.

We also report the PEA score among emotions in Figure 3. We observe that annotators agree with each other most often in *anger* and *fear*, with the average PEA scores at around 0.85. On the other hand, we find the least agreement in *surprise*. The PEA scores of the emotions are all above 0.66, indicating fairly high agreement among our annotators.

[6]To provide an intuitive interpretion, the PEA score of two adjacent primary emotions on the Plutchik's Wheel is 0.75.

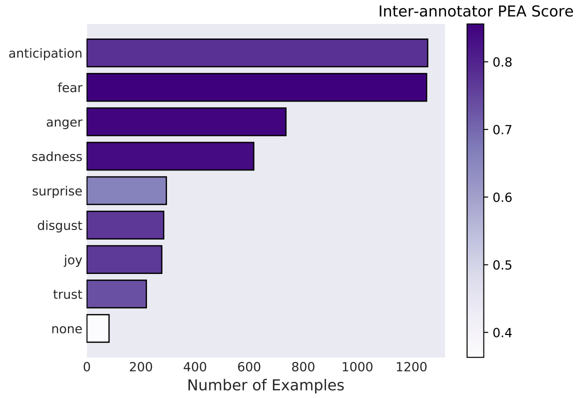

Figure 3: Emotion distribution of our dataset, ranked by the number of examples. Colors indicate the inter-annotator agreement measured by the PEA score.

Krippendorff's Alpha. In addition to the PEA metric, we also evaluate the inter-annotator agreement of emotion annotations in our dataset using the Krippendorff's Alpha Coefficient. We obtain an average value of 0.228 on all emotions using MASI distance.

4.2 Agreement in Triggers

Here we further examine the overlap in the annotated summaries of triggers when two annotators both select the same emotion for one example.

Extractive Summaries. We measure the percentage of overlap between two extractive summaries of trigger for the same emotion. We first pinpoint all the examples in which both annotators agree on the same emotion, then we extract the extractive summaries of trigger for the agreeing emotions. Next, we perform sentence tokenization on the extractive triggers and compute sentence-level overlap between them. We provide the detailed pre-processing procedure in Appendix §D.1.

As demonstrated in Table 1, within the posts where we find emotion overlap, 19.7% of the extractive summaries of trigger for the same emotion have completely identical annotations from both annotators, and 32.9% have partial sentence-level overlap. This suggests strong agreement between annotators.

Abstractive Summaries. We use Rouge to examine the overlap in abstractive triggers. Similarly, we first extract the abstractive summaries of trigger

Number of examples	1,889 Reddit posts
Number of emotions	8 primary emotions, each with 3 intensity labels
Number of unique annotators	15, including 2 expert linguists
Number of annotators per example	2 (either both turkers or both expert linguists)
Average number of emotions per example	2.62 ("None" excluded)
Number of examples with emotion agreement	1+ emotions agreement: 1546 (81.8%); 2+ emotions agreement: 531 (28.1%)
Percentage of overlapping extractive triggers for agreeing emotions	Complete overlap (sentence-level): 19.7%; Partial overlap (sentence-level): 32.9%
Average Rouge F1 score between abstractive triggers for agreeing emotions	Rouge-1: 0.256, Rouge-2: 0.056, Rouge-L: 0.191

Table 1: Summary statistics of our dataset.

for agreeing emotions, then we compute the Rouge score between them. As Table 1 shows, the average Rouge-1 F1 score between two annotators is 0.256, indicating distinctness in annotations on abstractive summaries of triggers.

4.3 Human Validation

In addition to the automatic evaluation metrics above, we also validate the emotion-trigger annotations in our dataset through human inspections. We set up a human validation task on the Amazon Mechanical Turk, and recruit a new group of qualified workers. We randomly sample 300 examples from our training set for validation. The emotion annotations, as well as both extractive and abstractive summaries of triggers, are validated.

We describe the validation framework as follows. The validators are given an annotated trigger summary. We first validate whether the summary actually indicates the annotated emotion by asking a yes/no question. Next, if the validator confirms the presence of emotion in the summary, we then ask whether the summary indeed expresses the *trigger* and not the *emotion* by raising another yes/no question. Extractive and abstractive summaries of triggers are validated separately. We present the validation results based on the extractive summaries in Table 2 and the results based on the abstractive summaries in Table 3. The numbers indicate the proportion of examples on which validators confirm upon.

Overall, the human validation results indicate fairly strong agreement in our annotations. The higher validated proportion in abstractive summaries than in extractive summaries can be explained by the sentence patterns commonly adopted by the annotators in abstractive summaries. For example, in expressing the abstractive trigger for *anger*, an annotation in our dataset is *I am angry that they would put me at risk of catching COVID and not tell me*, a sentence which is highly linguistically explicit of the emotion.

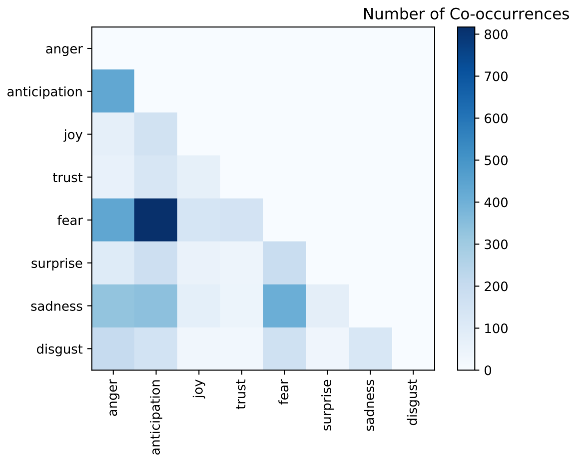

Figure 4: Emotion co-occurrence in our dataset.

5 Qualitative Analysis

5.1 Emotion Analysis

Emotion Distribution. Figure 3 shows the general emotion distribution of our dataset. *Anticipation* is the most common emotion across the dataset, closely followed by *fear*. There is clearly a huge gap among the emotions, with positively valenced emotions such as *trust* and *joy* rarely present in the dataset. This is predicted given the catastrophic nature of our domain.

Emotion Co-occurrence. We present the emotion co-occurrence heatmap in Figure 4. We observe that *anticipation* co-occurs with *fear* and *anger* most frequently in the dataset, suggesting

	Anticipation	Anger	Surprise	Fear	Sadness	Joy	Trust	Disgust	Avg
Emotion Presence	0.64	0.84	0.80	0.84	0.84	0.92	0.60	0.80	0.79
Trigger Presence	0.56	0.64	0.72	0.76	0.76	0.80	0.56	0.72	0.69

Table 2: Human validation results on the extractive summaries of triggers.

	Anticipation	Anger	Surprise	Fear	Sadness	Joy	Trust	Disgust	Avg
Emotion Presence	0.92	0.96	0.92	0.96	0.96	1.00	0.88	0.92	0.94
Trigger Presence	0.88	0.92	0.84	0.96	0.96	1.00	0.84	0.92	0.92

Table 3: Human validation results on the abstractive summaries of triggers.

Figure 5: Emotion distribution in our dataset over time (by week).

that Reddit posters are mostly anticipating negative events during COVID-19.

Fine-grained Intensity Levels. We convert the annotated intensity levels into scalar scores ranging from 1 to 3, with 3 indicating the most intense emotion. We report the average intensity scores of each emotion in Figure 6, with darker color indicating stronger intensity. We observe that *sadness* and *disgust* demonstrate the highest intensity level among all 8 emotions, whereas *anger* has the mildest intensity.

Emotion Trend. We present a time series analysis of the emotion distribution by week in Figure 5. The dataset is sampled from two timelines, with half of the posts dated during Summer 2021 and the other half dated since the outbreak of Omicron. In Figure 5, the two timelines are separated by a red line. We notice that the proportion of *anticipation* goes up consistently after Omicron, whereas *anger*, *sadness*, and *fear* drop down. This result is unsurprising since people are getting weary and tired after two years of avoiding COVID-19. It also provides evidence showing that Omicron is a milder disease.

Emotion Consistency. We examine how consistent the annotator is with the emotions in their extractive summaries of triggers. Our hypothesis is that similar sentences should be tagged with the same emotion. Here we measure the emotion consistency of one turker who is responsible for the majority of the annotations in the training and validation sets. We first extract all the extractive summaries along with the annotated emotions from the annotator. Then we select a prototypical extractive trigger for each emotion from the extractive summaries and use it as the anchor summary. Next, we use `all-mpnet-base-v2`, the top ranking model for Sentence Bert[7], to encode the sentences. We consider the top 5% extractive summaries closest in cosine similarity to each anchor summary the sentences that are semantically similar to the anchor summary. The sentences are manually validated by our expert linguists, who are asked to decide whether the sentences chosen based on cosine similarity to the anchor summary indeed express the target emotion. Both the preceding and exceeding sentences around the extractive trigger in the post are given to the validators as context information. After the validation, extractive summaries that do not express the target emotion are excluded.

We report the results averaged by the PEA score between the annotated emotion to the target emotion in Table 4. As the table shows, the annotator is fairly consistent in their emotion annotations.

5.2 Trigger Analysis

Linguistic Explicitness of Emotions. To examine the explicitness of emotions in the extractive triggers, we apply EmoLex (Mohammad and Turney, 2013), a dataset of English emotion associated lexicon human annotated on the Plutchik-8

[7] `https://www.sbert.net/docs/pretrained_models.html`

Emotions	Avg PEA Score
Anger	0.83
Fear	0.76
Anticipation	0.70
Surprise	0.51
Joy	0.80
Sadness	0.53
Trust	0.57
Disgust	0.53
Avg	0.65

Table 4: Emotion consistency results of one annotator in our dataset.

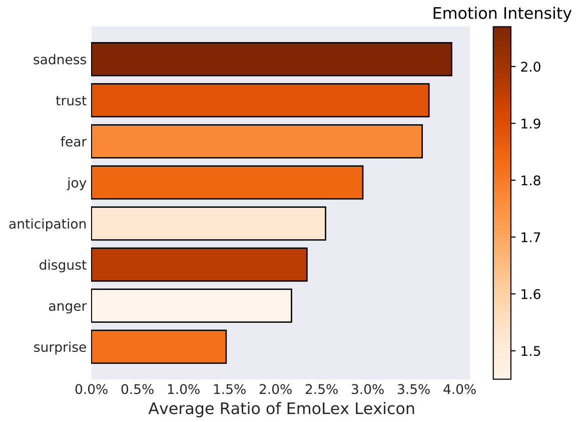

Figure 6: EmoLex lexicon ratio present in text. The color indicates the average intensity of the emotion in our dataset.

primary emotions. Specifically, for the extractive summaries of triggers to a certain emotion, we measure the average ratio of the said emotion words in EmoLex being present in the lemmatized summaries. The results are presented in Figure 6. We notice that *sadness* is the most explicit emotion in the annotated extractive summaries of triggers in our dataset, while *surprise* is the most implicit.

Extractive Trigger Position. We map the annotated extractive summaries of triggers into the original sentences in the post, and report the distribution of the triggers' positions in posts in Figure 7. Note that we do not repetitively include the post sentences: in cases where one sentence in the post contains multiple triggers, the sentence is only counted once in the distribution. We provide the detailed preprocessing procedure in Appendix §D.2. Results shows a large number of triggers in the first sentences of the post.

Abstractive Trigger Abstractiveness. We measure the abstractiveness of the annotated abstractive summaries of triggers by computing the Rouge score between the extractive and abstractive summaries. we use ROUGE-n precision scores to calculate how abstractive the annotated abstractive summaries are compared to the annotated extractive summaries. Results are: Rouge-1: 0.27, Rouge-2: 0.072, Rouge-L: 0.206. The results indicate that the abstractive summaries are fairly abstract from the extractive summaries in our dataset.

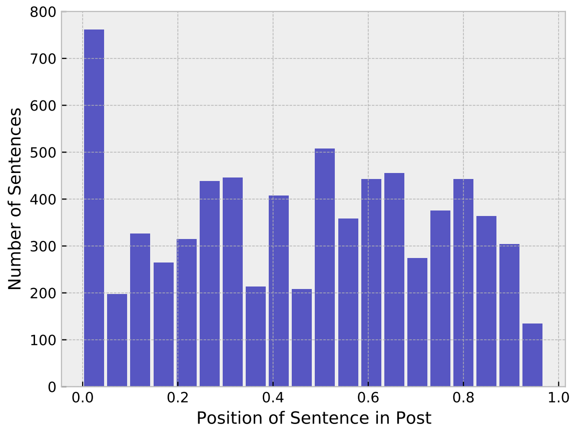

Figure 7: Extractive triggers' distribution in the original posts. 0 means the first sentence of the post, and 1 means the last sentence.

Topic Variation. To better understand the triggers of each emotion, we use the Named-Entity Recognition (NER) to extract frequent entities in the extractive triggers. Results are shown in Figure 8.

6 Baseline Modeling

In this section, we benchmark several methods on our dataset. Using our dataset, we evaluate the ability to which large-scale pre-trained language models transfer from existing resources. We assess state-of-the-art models' ability on three tasks respectively, namely emotion detection, emotion-based extractive trigger summarization, and emotion-based abstractive trigger summarization. Note that given the novelty of the Emotion-Trigger Summarization task, none of the current summarization methods are emotion-based. In other words, the models would produce the same summaries for different emotions in a document.

The structure of this section is as follows. We briefly introduce the existing emotion detection and automatic text summarization datasets in §6.1. Then we describe our experiment setup together

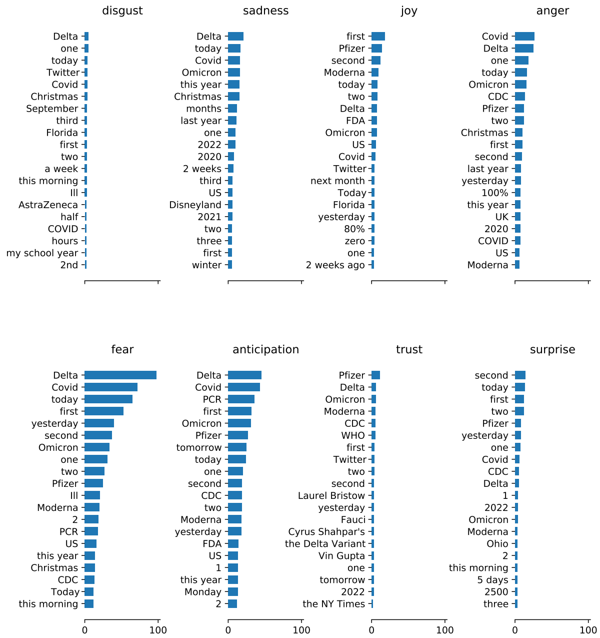

Figure 8: NER results of annotated extractive triggers, ranked by number of occurrences.

with the state-of-the-art models we are assessing in §6.2. Finally, we report the results in §6.3.

6.1 Existing Datasets

Here we introduce existing datasets on the tasks of emotion detection as well as automatic text summarization. All the datasets are in English.

Emotion Detection. *GoEmotions* (Demszky et al., 2020) is an emotion dataset from general Reddit comments, in which every example is annotated with one or more than one of 27 emotions or Neutral. Similar to our dataset, GoEmotions is also sourced from Reddit. However, the examples in GoEmotions are comment sentences from a large number of subreddits. Using GoEmotions, we evaluate how well emotion detection models trained outside the COVID-19 domain generalize to our dataset. Following Sosea et al. (2021), we map GoEmotions labels into Plutchik-8 emotions[8].

Text Summarization. *CNN/DailyMail* (Hermann et al., 2015) is a text summarization dataset consisting of news articles and associated highlights. The highlights depict a brief overview of the article.

XSum (Narayan et al., 2018) is a dataset of news articles, where each article is accompanied with a highly abstractive one-sentence summary, answering the question "What is this article about?".

Webis-TLDR-17 (Völske et al., 2017) is an abstractive text summarization dataset on the social media Reddit. Unlike the two news datasets above, Webis-TLDR-17 is closer in domain to our dataset.

6.2 Experimental Setup

We use the Hugging Face framework (Wolf et al., 2020) for reproductivity in the following models. Hyperparameters used for the best models are presented in Appendix §E.

Emotion Detection. We first employ a lexical baseline to detect emotions using *EmoLex* (Mohammad and Turney, 2013), a word-emotion lexicon on the 8 Plutchik primary emotions.

We also explore how well emotion detection models trained outside our domain generalize to the COVID context. We experiment with the `bert-base-uncased` (Devlin et al., 2019) model and fine-tune it on the GoEmotions dataset.

Emotion-based Trigger Summarization. We evaluate the following models on our dataset: **1)** `BART` (Lewis et al., 2020); **2)** `Pegasus` (Zhang et al., 2019); **3)** `PreSumm` (Liu and Lapata, 2019); and **4)** `PacSum` (Zheng and Lapata, 2019). We fine-tune these models on the CNN/DailyMail, XSum, and Webis-TLDR-17 corpuses respectively and report their performance on our benchmark.

[8]GoEmotions Mapping: Anger → Anger; Disgust → Disgust; Joy → Joy; Sadness → Sadness; Fear → Fear; Nervousness, Desire → Anticipation; Surprise → Surprise; Admiration → Trust.

	Anger	Sadness	Joy	Fear	Surprise	Anticipation	Trust	Disgust	AVG
EmoLex	0.356	0.425	0.487	0.567	0.030	0.178	0.135	0.205	0.298

Table 5: Results of baseline models on emotion detection.

Anger	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.211	0.131	0.190	0.151	0.030	0.121
3 Sentence	0.341	0.254	0.307	0.209	0.045	0.142
EmoSentence	0.215	0.214	0.254	0.164	0.027	0.111
Pegasus-Dailymail	0.346/0.393	0.262/0.298	0.314/0.361	0.208/0.266	0.043/0.088	0.142/0.216
Pegasus-XSum	0.156/0.298	0.037/0.191	0.129/0.265	0.154/0.252	0.024/0.097	0.129/0.225
Pegasus-Reddit	0.241/0.391	0.128/0.295	0.202/0.359	0.188/0.251	0.044/0.088	0.138/0.209
BART-large-CNN/Dailymail	0.396/0.404	0.300/0.306	0.348/0.368	0.272/0.318	0.062/0.110	0.169/0.219
BART-large-XSum	0.193/0.286	0.066/0.199	0.158/0.269	0.172/0.308	0.026/0.119	0.134/0.244
(Abst) BertSumExtAbs-CNN/Dailymail	0.308	0.203	0.269	0.230	0.051	0.157
(Abst) BertSumExtAbs-XSum	0.129	0.018	0.103	0.146	0.013	0.112
(Ext) BertSumExt-CNN/Dailymail	0.357	0.261	0.312	0.240	0.049	0.153
PacSum	0.389	0.294	0.344	0.260	0.059	0.164
Anticipation	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.183	0.091	0.163	0.150	0.026	0.119
3 Sentence	0.287	0.188	0.250	0.207	0.041	0.142
EmoSentence	0.181	0.086	0.159	0.157	0.021	0.117
Pegasus-Dailymail	0.342/0.422	0.259/0.338	0.312/0.395	0.221/0.307	0.049/0.109	0.152/0.251
Pegasus-XSum	0.198/0.408	0.080/0.317	0.173/0.384	0.185/0.284	0.035/0.109	0.151/0.247
Pegasus-Reddit	0.252/0.418	0.139/0.335	0.214/0.391	0.196/0.293	0.046/0.108	0.144/0.245
BART-large-CNN/Dailymail	0.353/0.423	0.253/0.337	0.302/0.380	0.249/0.344	0.050/0.125	0.159/0.243
BART-large-XSum	0.193/0.353	0.075/0.285	0.160/0.340	0.193/0.310	0.043/0.127	0.148/0.256
(Abst) BertSumExtAbs-CNN/Dailymail	0.250	0.139	0.217	0.213	0.040	0.151
(Abst) BertSumExtAbs-XSum	0.126	0.014	0.100	0.149	0.019	0.115
(Ext) BertSumExt-CNN/Dailymail	0.325	0.226	0.275	0.228	0.042	0.147
PacSum	0.311	0.202	0.261	0.235	0.044	0.151
Joy	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.198	0.139	0.190	0.126	0.020	0.103
3 Sentence	0.301	0.231	0.279	0.208	0.055	0.154
EmoSentence	0.215	0.132	0.195	0.124	0.024	0.105
Pegasus-Dailymail	0.290/0.334	0.212/0.270	0.265/0.318	0.206/0.213	0.051/0.061	0.150/0.178
Pegasus-XSum	0.145/0.267	0.039/0.189	0.131/0.250	0.129/0.152	0.018/0.027	0.114/0.136
Pegasus-Reddit	0.165/0.245	0.079/0.184	0.141/0.232	0.156/0.148	0.034/0.032	0.124/0.119
BART-large-CNN/Dailymail	0.382/0.368	0.316/0.284	0.359/0.339	0.205/0.248	0.045/0.066	0.143/0.178
BART-large-XSum	0.231/0.309	0.114/0.242	0.198/0.303	0.158/0.195	0.019/0.057	0.122/0.172
(Abst) BertSumExtAbs-CNN/Dailymail	0.384	0.293	0.354	0.191	0.034	0.142
(Abst) BertSumExtAbs-XSum	0.143	0.017	0.120	0.141	0.011	0.107
(Ext) BertSumExt-CNN/Dailymail	0.318	0.230	0.289	0.184	0.040	0.125
PacSum	0.363	0.279	0.324	0.193	0.039	0.133
Trust	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.143	0.050	0.118	0.121	0.016	0.094
3 Sentence	0.199	0.076	0.152	0.171	0.027	0.117
EmoSentence	0.141	0.040	0.113	0.116	0.017	0.092
Pegasus-Dailymail	0.301/0.243	0.210/0.130	0.272/0.207	0.188/0.201	0.030/0.035	0.134/0.140
Pegasus-XSum	0.151/0.210	0.052/0.094	0.132/0.189	0.158/0.133	0.007/0.014	0.130/0.124
Pegasus-Reddit	0.231/0.296	0.131/0.216	0.204/0.270	0.126/0.143	0.019/0.026	0.096/0.114
BART-large-CNN/Dailymail	0.318/0.345	0.212/0.247	0.275/0.287	0.172/0.224	0.037/0.073	0.130/0.182
BART-large-XSum	0.166/0.216	0.030/0.121	0.131/0.187	0.170/0.249	0.016/0.069	0.147/0.217
(Abst) BertSumExtAbs-CNN/Dailymail	0.311	0.188	0.259	0.186	0.029	0.139
(Abst) BertSumExtAbs-XSum	0.156	0.024	0.115	0.140	0.009	0.111
(Ext) BertSumExt-CNN/Dailymail	0.255	0.148	0.207	0.171	0.033	0.121
PacSum	0.275	0.171	0.243	0.169	0.033	0.116

Table 6: Results of baseline models on emotion-based trigger summarization.

Fear	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.218	0.128	0.198	0.155	0.028	0.122
3 Sentence	0.312	0.222	0.282	0.222	0.495	0.153
EmoSentence	0.225	0.123	0.197	0.157	0.023	0.125
Pegasus-Dailymail	0.357/0.410	0.278/0.316	0.330/0.376	0.234/0.316	0.054/0.110	0.158/0.250
Pegasus-XSum	0.167/0.359	0.050/0.250	0.141/0.323	0.171/0.286	0.035/0.105	0.140/0.242
Pegasus-Reddit	0.255/0.410	0.130/0.315	0.208/0.376	0.214/0.313	0.053/0.109	0.156/0.248
BART-large-CNN/Dailymail	0.394/0.460	0.290/0.372	0.340/0.415	0.270/0.349	0.062/0.125	0.168/0.242
BART-large-XSum	0.194/0.296	0.074/0.217	0.163/0.280	0.187/0.320	0.040/0.129	0.140/0.248
(Abst) BertSumExtAbs-CNN/Dailymail	0.280	0.159	0.232	0.218	0.043	0.151
(Abst) BertSumExtAbs-XSum	0.133	0.017	0.107	0.144	0.016	0.106
(Ext) BertSumExt-CNN/Dailymail	0.345	0.230	0.294	0.247	0.055	0.155
PacSum	0.367	0.262	0.321	0.253	0.057	0.161
Surprise	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.289	0.200	0.270	0.181	0.039	0.143
3 Sentence	0.354	0.283	0.332	0.230	0.062	0.165
EmoSentence	0.281	0.254	0.267	0.176	0.037	0.141
Pegasus-Dailymail	0.330/0.334	0.253/0.241	0.306/0.311	0.238/0.222	0.065/0.060	0.174/0.167
Pegasus-XSum	0.154/0.277	0.037/0.188	0.122/0.254	0.176/0.199	0.034/0.051	0.440/0.162
Pegasus-Reddit	0.253/0.344	0.117/0.254	0.202/0.303	0.223/0.201	0.061/0.542	0.160/0.155
BART-large-CNN/Dailymail	0.397/0.376	0.323/0.285	0.360/0.325	0.238/0.289	0.059/0.083	0.152/0.208
BART-large-XSum	0.206/0.322	0.083/0.227	0.173/0.296	0.202/0.287	0.041/0.088	0.150/0.240
(Abst) BertSumExtAbs-CNN/Dailymail	0.309	0.197	0.269	0.237	0.058	0.160
(Abst) BertSumExtAbs-XSum	0.134	0.015	0.108	0.146	0.017	0.118
(Ext) BertSumExt-CNN/Dailymail	0.351	0.255	0.314	0.239	0.056	0.160
PacSum	0.353	0.252	0.306	0.236	0.055	0.157
Sadness	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.188	0.091	0.168	0.140	0.022	0.115
3 Sentence	0.281	0.192	0.250	0.193	0.035	0.134
EmoSentence	0.193	0.094	0.167	0.145	0.025	0.118
Pegasus-Dailymail	0.332/0.349	0.239/0.247	0.297/0.318	0.210/0.263	0.045/0.083	0.144/0.204
Pegasus-XSum	0.149/0.355	0.032/0.248	0.124/0.328	0.149/0.250	0.027/0.092	0.123/0.218
Pegasus-Reddit	0.248/0.350	0.135/0.252	0.210/0.317	0.184/0.263	0.040/0.088	0.136/0.210
BART-large-CNN/Dailymail	0.415/0.472	0.324/0.389	0.355/0.427	0.250/0.325	0.055/0.111	0.157/0.226
BART-large-XSum	0.196/0.296	0.081/0.227	0.169/0.283	0.173/0.282	0.032/0.101	0.134/0.226
(Abst) BertSumExtAbs-CNN/Dailymail	0.300	0.181	0.260	0.211	0.037	0.147
(Abst) BertSumExtAbs-XSum	0.161	0.026	0.123	0.155	0.013	0.121
(Ext) BertSumExt-CNN/Dailymail	0.331	0.227	0.280	0.220	0.042	0.137
PacSum	0.325	0.209	0.275	0.217	0.037	0.140
Disgust	Extractive			Abstractive		
	Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
First Sentence	0.163	0.076	0.144	0.135	0.015	0.112
3 Sentence	0.285	0.195	0.253	0.182	0.029	0.129
EmoSentence	0.167	0.075	0.147	0.132	0.017	0.111
Pegasus-Dailymail	0.290/0.304	0.188/0.202	0.252/0.270	0.176/0.182	0.027/0.026	0.126/0.134
Pegasus-XSum	0.189/0.276	0.081/0.182	0.173/0.260	0.159/0.244	0.012/0.058	0.127/0.198
Pegasus-Reddit	0.201/0.308	0.094/0.219	0.171/0.277	0.158/0.175	0.033/0.037	0.126/0.132
BART-large-CNN/Dailymail	0.323/0.341	0.229/0.240	0.282/0.301	0.211/0.273	0.036/0.077	0.140/0.193
BART-large-XSum	0.197/0.291	0.074/0.198	0.167/0.274	0.160/0.253	0.027/0.083	0.123/0.218
(Abst) BertSumExtAbs-CNN/Dailymail	0.313	0.207	0.284	0.202	0.035	0.151
(Abst) BertSumExtAbs-XSum	0.116	0.008	0.091	0.126	0.010	0.095
(Ext) BertSumExt-CNN/Dailymail	0.299	0.193	0.262	0.208	0.038	0.140
PacSum	0.311	0.214	0.261	0.217	0.036	0.142

Table 7: (Continued) Results of baseline models on emotion-based trigger summarization.

6.3 Results

The emotion detection performance is listed in Table 5. The emotion-based summarization results are presented in Table 6 and Table 7. Performance of the models that are further fine-tuned on our dataset is presented on the right of the slash. Results show that fine-tuning the Pegasus and BART models on our dataset improves summarization performance in terms of ROUGE. In conclusion, our benchmark dataset is able to successfully address the current task of Emotion-Trigger Summarization.

7 Future Work: Task-Guided Pre-Training

In the future work, we aim to incorporate a joint model that includes task-guided pre-training on Emotion-Trigger Summarization. We notice that the number of posts in the subreddit r/-COVID19_support is very limited even without length filtering. Therefore, for pre-training purposes, we turn to r/COVID-19Positive[9], a subreddit where people who tested positive for COVID-19 share their experiences. We manually inspect the subreddits related to COVID-19 and find that posts on r/COVID-19Positive are most similar to those on r/COVID19_support.

8 Acknowledgements

I would like to thank Dr. Jessy Li, Dr. Cornelia Caragea, Tiberiu Sosea, and John Henry Cruz for their inputs into this project. I would also like to thank our annotators for their hard work.

[9] https://www.reddit.com/r/COVID19positive/

References

Muhammad Abdul-Mageed and Lyle Ungar. 2017. EmoNet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 718–728, Vancouver, Canada. Association for Computational Linguistics.

Mehdi Allahyari, Seyed Amin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, and Krys J. Kochut. 2017. Text summarization techniques: A brief survey. *CoRR*, abs/1707.02268.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579–586, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Ashwini Ashokkumar and James W Pennebaker. 2021. Social media conversations reveal large psychological shifts caused by COVID-19's onset across U.S. cities. *Science advances*, 7(39):eabg7843–eabg7843. Edition: 2021/09/22 Publisher: American Association for the Advancement of Science.

Fang Chen, Ziwei Shi, Zhongliang Yang, and Yongfeng Huang. 2022. Recurrent synchronization network for emotion-cause pair extraction. *Knowledge-Based Systems*, 238:107965.

Ying Chen, Wenjun Hou, Xiyao Cheng, and Shoushan Li. 2018. Joint learning for emotion classification and emotion cause detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 646–651, Brussels, Belgium. Association for Computational Linguistics.

Ying Chen, Wenjun Hou, Shoushan Li, Caicong Wu, and Xiaoqiang Zhang. 2020. End-to-end emotion-cause pair extraction with graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 198–207, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 179–187, Beijing, China. Coling 2010 Organizing Committee.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Shrey Desai, Cornelia Caragea, and Junyi Jessy Li. 2020. Detecting perceived emotions in hurricane disasters. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5290–5305, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020. ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3161–3170, Online. Association for Computational Linguistics.

Brett Drury, Hugo Gonçalo Oliveira, and Alneu Lopes. 2022. A survey of the extraction and applications of causal relations. *Natural Language Engineering*, pages 1–40.

Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea, and Hoda K. Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165:113679.

Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Lidong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. A knowledge regularized hierarchical approach for emotion cause analysis. 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)*, pages 5614–5624, Hong Kong, China. Association for Computational Linguistics.

Chuang Fan, Chaofa Yuan, Jiachen Du, Lin Gui, Min Yang, and Ruifeng Xu. 2020. Transition-based directed graph construction for emotion-cause pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3707–3717, Online. Association for Computational Linguistics.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.

Wei Gao, Shoushan Li, Sophia Yat Mei Lee, Guodong Zhou, and Chu-Ren Huang. 2013. Joint learning on sentiment and emotion classification. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, CIKM '13, page 1505–1508, New York, NY, USA. Association for Computing Machinery.

Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In *CICLing*.

Matej Gjurković and Jan Šnajder. 2018. Reddit: A gold mine for personality prediction. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 87–97, New Orleans, Louisiana, USA. Association for Computational Linguistics.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach for emotion cause extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1593–1602, Copenhagen, Denmark. Association for Computational Linguistics.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-driven emotion cause extraction with corpus construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1639–1649, Austin, Texas. Association for Computational Linguistics.

Hiroaki Hayashi, Prashant Budania, Peng Wang, Chris Ackerson, Raj Neervannan, and Graham Neubig. 2021. WikiAsp: A dataset for multi-domain aspect-based summarization. *Transactions of the Association for Computational Linguistics*, 9:211–225.

Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems (NIPS)*.

Jiaxing Hu, Shumin Shi, and Heyan Huang. 2019. Combining external sentiment knowledge for emotion cause detection. In *Natural Language Processing and Chinese Computing*, pages 711–722, Cham. Springer International Publishing.

Zhengping Jiang, Sarah Ita Levitan, Jonathan Zomick, and Julia Hirschberg. 2020. Detection of mental health from Reddit via deep contextualized representations. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pages 147–156, Online. Association for Computational Linguistics.

Hamed Khanpour and Cornelia Caragea. 2018. Fine-grained emotion detection in health-related online posts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1160–1166, Brussels, Belgium. Association for Computational Linguistics.

Arunima Khunteta and Pardeep Singh. 2021. Emotion cause extraction - a review of various methods and corpora. In *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*, pages 314–319.

Sophia Yat Mei Lee, Ying Chen, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause events: Corpus construction and analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Jingsheng Lei, Yanghui Rao, Qing Li, Xiaojun Quan, and Liu Wenyin. 2014. Towards building a social emotion detection system for online news. *Future Generation Computer Systems*, 37:438–448.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Shoushan Li, Lei Huang, Rong Wang, and Guodong Zhou. 2015. Sentence-level emotion classification with label and context dependence. In *ACL*.

Weiyuan Li and Hua Xu. 2014. Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 41(4, Part 2):1742–1749.

Xiangju Li, Kaisong Song, Shi Feng, Daling Wang, and Yifei Zhang. 2018. A co-attention neural network model for emotion cause analysis with emotional context awareness. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4752–4757, Brussels, Belgium. Association for Computational Linguistics.

Chen Liu, Muhammad Osama, and Anderson De Andrade. 2019. DENS: A dataset for multi-class emotion analysis. 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)*, pages 6293–6298, Hong Kong, China. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. 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)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

Rada Mihalcea and Carlo Strapparava. 2012. Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 590–599, Jeju Island, Korea. Association for Computational Linguistics.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

M. F. Mridha, Aklima Akter Lima, Kamruddin Nur, Sujoy Chandra Das, Mahmud Hasan, and Muhammad Mohsin Kabir. 2021. A survey of automatic text summarization: Progress, process and challenges. *IEEE Access*, 9:156043–156070.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Alena Neviarouskaya and Masaki Aono. 2013. Extracting causes of emotions from text. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 932–936, Nagoya, Japan. Asian Federation of Natural Language Processing.

Inna Pirina and Çağrı Çöltekin. 2018. Identifying depression on Reddit: The effect of training data. In *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task*, pages 9–12, Brussels, Belgium. Association for Computational Linguistics.

Robert Plutchik. 1991. *The Emotions*. University Press of America.

Robert Plutchik. 2001. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350.

Irene Russo, Tommaso Caselli, Francesco Rubino, Ester Boldrini, and Patricio Martínez-Barco. 2011. EMOCause: An easy-adaptable approach to extract emotion cause contexts. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 153–160, Portland, Oregon. Association for Computational Linguistics.

Sarah Seraj, Kate G. Blackburn, and James W. Pennebaker. 2021. Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7):e2017154118.

Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through Reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, pages 58–65, Vancouver, BC. Association for Computational Linguistics.

Tiberiu Sosea and Cornelia Caragea. 2020. CancerEmo: A dataset for fine-grained emotion detection. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8892–8904, Online. Association for Computational Linguistics.

Tiberiu Sosea, Chau Pham, Alexander Tekle, Cornelia Caragea, and Junyi Jessy Li. 2021. Emotion analysis and detection during covid-19.

Carlo Strapparava and Rada Mihalcea. 2007. SemEval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74, Prague, Czech Republic. Association for Computational Linguistics.

Bowen Tan, Lianhui Qin, Eric Xing, and Zhiting Hu. 2020. Summarizing text on any aspects: A knowledge-informed weakly-supervised approach. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6301–6309, Online. Association for Computational Linguistics.

Elsbeth Turcan and Kathy McKeown. 2019. Dreaddit: A Reddit dataset for stress analysis in social media. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 97–107, Hong Kong. Association for Computational Linguistics.

Michael Völske, Martin Potthast, Shahbaz Syed, and Benno Stein. 2017. TL;DR: Mining Reddit to learn automatic summarization. In *Proceedings of the Workshop on New Frontiers in Summarization*, pages 59–63, Copenhagen, Denmark. Association for Computational Linguistics.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and A. Sheth. 2012. Harnessing twitter "big data" for automatic emotion identification. *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 587–592.

Zhaoxia Wang, Seng-Beng Ho, and E. Cambria. 2020. A review of emotion sensing: categorization models and algorithms. *Multimedia Tools and Applications*, pages 1–30.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Rui Xia and Zixiang Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. *CoRR*, abs/1906.01267.

Xinglin Xiao, Penghui Wei, Wenji Mao, and Lei Wang. 2019. Context-aware multi-view attention networks for emotion cause extraction. In *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 128–133. IEEE.

Bo Xu, Hongfei Lin, Yuan Lin, Yufeng Diao, Liang Yang, and Kan Xu. 2019. Extracting emotion causes using learning to rank methods from an information retrieval perspective. *IEEE Access*, 7:15573–15583.

Jun Xu, Ruifeng Xu, Qin Lu, and Xiaolong Wang. 2012. Coarse-to-fine sentence-level emotion classification based on the intra-sentence features and sentential context. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, page 2455–2458, New York, NY, USA. Association for Computing Machinery.

Ruifeng Xu, Jiannan Hu, Qin Lu, Dongyin Wu, and Lin Gui. 2017. An ensemble approach for emotion cause detection with event extraction and multi-kernel svms. *Tsinghua Science and Technology*, 22(6):646–659.

Shuntaro Yada, K. Ikeda, Keiichiro Hoashi, and Kyo Kageura. 2017. A bootstrap method for automatic rule acquisition on emotion cause extraction. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 414–421.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2019. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization.

Hao Zheng and Mirella Lapata. 2019. Sentence centrality revisited for unsupervised summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6236–6247, Florence, Italy. Association for Computational Linguistics.

Ming Zhong, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, and Dragomir Radev. 2021. QMSum: A new benchmark for query-based multi-domain meeting summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5905–5921, Online. Association for Computational Linguistics.

Appendix A Data Curation Details

Here we detail the preprocessing procedure over the source data. We preprocess the source data using regular expressions. As the first step, we tokenize posts into individual words. Specifically, we apply the following regular expressions in combination with the NLTK word_tokenize package to tokenize posts into words:

```
re.sub("\s+"," ", post)
re.sub(r'(?<=[.,!?:])(?=[^\s])', r' ',
    post)
re.sub(r'\s([?.!,:"](?:\s|$))', r'\1',
    post)
nltk.word_tokenize(post)
```

Then we exclude punctuation from the tokenized posts and filter the posts that are 50-400 tokens long. Finally, we mask web links by substituting them into `[url]` tokens using the following regular expression:

```
replace(r'http\S+', '[url]').str.strip()
replace(r'''(?i)\b((?:https?://|www\d
    {0,3}[.]|[a-z0-9.\-]+[.][a-z]{2,4}/)
    (?:[^\s()<>]+|\(([^\s()<>]+|(\([^\s()
    <>]+\)))*\))+(?:\(([^\s()<>]+|(\([^\
    s()<>]+\)))*\)|[^\s`!()
    \[\]{};:'".,<>?«»“”‘’]))''', '[url]')
    .str.strip()
```

Appendix B Annotation Details

B.1 Annotators' Personality Surveys

We distribute two personality surveys to our annotators, namely the *Ten-Item Personality Inventory (TIPI) of the Big Five personality traits* and the *Interpersonal Reactivity Index*. Permission of consent is endowed by the annotators. We have collected the surveys from five major annotators on our dataset. Results are listed in Table 8.

B.2 Annotation Instructions

Comprehensive instructions are provided to the annotators, as demonstrated in Figure B.9. Note that the instruction page pops up as a modal before every annotation, so as to remind the annotators of the task framework. We also ask the annotators to pay special attention to a few principles as follows. For the emotion annotations, we ask annotators to follow the emotion guidelines on the Six Seconds website[10] and interpret *anticipation* as (good or bad) *expectancy* (Plutchik, 1991). For the trigger annotations, we instruct annotators to annotate summaries containing *triggers* that lead to the emotion instead of sentences expressing the *emotion* itself.

The layout of our annotation task is shown in Figure B.10.

[10] https://www.6seconds.org/2020/08/11/plutchik-wheel-emotions/

B.3 Pre-Annotation Training

We pre-train the annotators before they annotate our dataset. We set up a qualification task on the Amazon Mechanical Turk. The qualification task involves 3 posts, and annotators are required to complete the qualification task. Through manually examining the annotators' work on the qualification task and comparing the annotations to the *gold* annotations we make, we filter high-quality annotators and give them the access to our annotation task. We also provide feedback to their annotations.

The turkers are paid a minimal wage of $10 per hour. To ensure this goal is reached, we keep track of their working time on the backstage and give out bonuses accordingly when needed.

B.4 Annotation Revisions

During the process of the annotation on our dataset, we regularly review the annotations and give feedback accordingly. When needed, we send the annotations back to the annotator along with the original post, and ask them to revise their work based on our suggestions. Note that the annotator is responsible for the revision of his own work only.

Appendix C Dataset Examples

An example of our dataset is shown in Figure C.11. This example includes annotations from both annotators. Annotations for different emotions are in distinct colors.

Appendix D Annotation Preprocessing

D.1 Extractive Trigger Overlap

To measure the overlap between extractive triggers for the same emotion in a post, we first tokenize the annotations into sentences. Even though annotators are instructed to cut and paste the original clauses/sentences in the post, there are some irregularities in the annotations that impede sentence tokenizations. For example, annotators use various special tokens including <sep>, \, and ... to separate multiple extractive summaries of triggers if they come from different places of the post. In order to cleanly split the extractive summaries into sentences and measure sentence-level overlap

	Interpersonal Reactivity Index		Ten-Item Personality Inventory				
Annotators	Empathic Concern	Perspective Taking	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experiences
A	24	16	11	9	11	8	8
B	17	20	9	8	11	10	12
C	21	19	4	10	12	5	10
D	24	28	10	13	14	14	7
E	3	0	3	2	2	2	2
Mean	17.8	16.6	7.4	8.4	10	7.8	7.8
SD	7.83326	9.2	3.2619	3.6111	4.1473	4.1183	3.3705

Table 8: Survey results.

between two annotators, we apply the following regular expressions:

```
nltk.sent_tokenize(ext_trigger)
re.split(" / ", sent, flags=re.
    IGNORECASE)
re.split("<sep>", sent, flags=re.
    IGNORECASE)
re.split("\n\n", sent)
re.split("\n", sent)
re.sub(r'[^\w\s]', '', sent)
```

Finally, we strip and lower the sentences.

D.2 Extractive Summaries of Trigger Position: Preprocessing Details

To measure the position of the annotated extractive summaries of trigger in the original post, we first tokenize the annotations into individual sentences. The following codes are used:

```
re.sub(r'(?<=[.,!?:])(?=[^\s])', r" ",
    ext_trigger)
nltk.sent_tokenize(ext_trigger)
re.split(" / ", sent, flags=re.
    IGNORECASE)
re.split("<sep>", sent, flags=re.
    IGNORECASE)
re.split("\n\n", sent)
re.split("\n", sent)
re.sub(r'[^\w\s]', '', sent)
```

We also tokenize the posts into sentences in lower case. We identify the original post sentence containing the extractive trigger. In cases where the extractive trigger sentence cannot be found in the original post (for example, the annotator purposefully ellipses an uninformative part in the middle of the original sentence), we compute the Rouge-2 precision score of the extractive trigger sentence to each sentence in the original post, and use the post sentence with the highest score as the target sentence.

Appendix E Model Hyperparameters

Here we provide the hyperparameters used in our best models.

×

Help us understand COVID-19 Emotions!

- Each of the displayed texts represents a Reddit post related to COVID-19.
- Please select the displayed emotion(s).
- After selecting the emotion(s), provide extractive and abstractive summaries of the triggers.
- The length of your summary should be limited to one or two sentences. Use commas if there are multiple parts.
- **Extractive**: find **clauses/sentences** in the Reddit post that best summarizes **what causes the emotion**.
- **Abstractive**: if you feel like summarizing **what causes the emotion** using you own words, feel free to do so. Summarize in the perspective of the poster.
- **Note**: A text can still contain an emotion even if there is no explicit emotion cue.
- **Note**: A text may display multiple emotions and more than one emotion can be selected.
- **Note**: For **"Anticipation"**, you can anticipate both good and bad things.
- **Note**: This website will help you understand the emotions better, please click here.

Please read the following examples:

Reddit: "COVID has been around for two years now. I've been reading the news about COVID conditions since the pandemic struck. Just today I saw lower case numbers and the rate of transmission falling below 0.9. A large decline in the number of new weekly deaths was also reported for all regions. For a long time I feared that this day would never come and that Delta was going to always keep on spreading. What a relief to find out that I was wrong."
Emotion(s):
- **Surprise**:
Extractive: Just today I saw lower case numbers and the rate of transmission falling below 0.9.
Abstractive: Improving conditions of COVID in my area is a nice surprise.
- **Joy**:
Extractive: For a long time I feared that this day would never come and that Delta was going to always keep on spreading. What a relief to find out that I was wrong.
Abstractive: I am happy because I found out that the COVID conditions are getting better.

Below are some example sentences of the emotions:

1. Anger:
1) I feel angry. I up and MOVED across the country to get away from anti-vaxxers, why would I knowingly hang out with an unvaccinated person?
2) Let me get this straight: you recognize that it would "stifle economic activity" but would rather these construction companies and their workers halt operations, NOT because of their health and safety due to COVID19, but, instead, so your 2-year-old can take a fucking nap while you WORK FROM HOME?

2. Anticipation:
1) Live music is our great connector though and I really hope it'd help me feel less disconnected.
2) After this wave maybe we can start to hope a little bit more, between July and November Pfizer is supposed to deliver 186 million doses enough to fully vax more than half of the population in 5 months so I guess that we could see a somewhat Quick end to this pandemic if everything goes right with delivery.

3. Joy:
1) I'm excited, this feels like it will be a big step towards getting back to my normal life after all this time. I just wanted to say this because I'm happy about it.
2) This is great seeing business take matter into their own hands and opening back up!

4. Trust:
1) This is the article I've found to be most concise in explaining everything.
2) So long as people take the vaccines, ending coronavirus shouldn't be a concern.

5. Fear:
1) I don't feel like going to the bar because I'm afraid of catching COVID.
2) I just had a dentist visit and all COVID precautions were followed, but I am still experiencing anxiety over potential exposure.

6. Surprise:
1) I'm shocked to find out that the city I just arrived at apparently ran into some controversy since they flew some infected people in recently.
2) Nail salons and bowling alleys are reopen for business now. I don't see where in the governor's order or the Texas DSHS guidelines that they are allowed to open.

7. Sadness:
1) My nanny got sick and hasn't been out for the past 2.5 weeks. She calls me crying, saying she has panic, shortness of breath, chest pain, and many other symptoms. I feel overwhelmed, sad, and bad for her.
2) My depression has gotten so bad because I've been in the house for so long. This pandemic having no end in sight, people dying left and right, the constant fear mongering, life being so dull and the future uncertain... I hate not being able to do anything.

8. Disgust:
1) Yikes, people are overreacting to COVID. It's not some super deadly virus, it all comes down to the individual's immune response and what kind of treatment is available for their symptoms.
2) Reopening the state is a purely political move from Kemp. Kemp and the GOP work together with big businesses. The big businesses need their slaves – I mean workers to make them money. Closed economy means no money flowing into their pockets.

Close

Figure B.9: Annotation instructions (always shown before annotating).

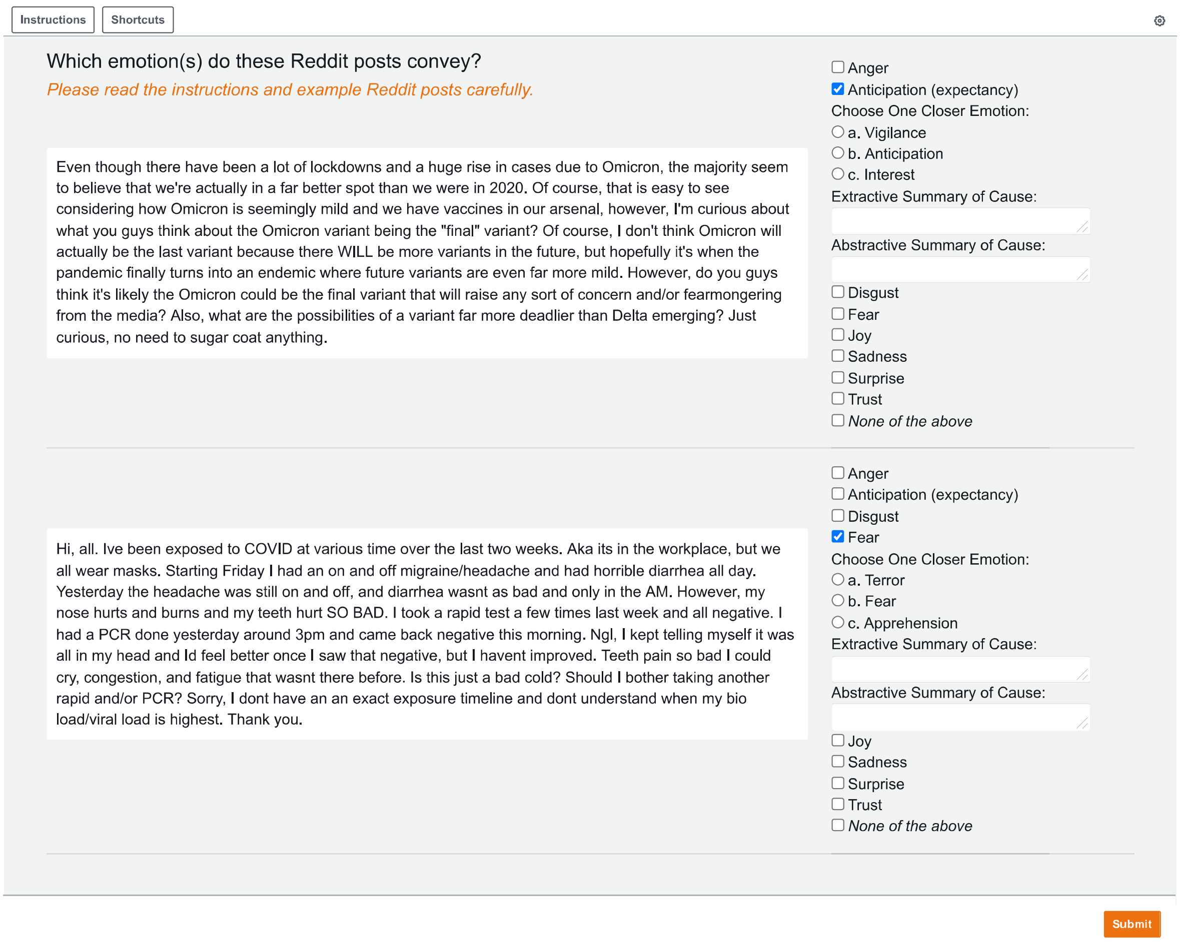

Figure B.10: The annotation task layout of an example hit.

Reddit Post

Hi everyone, I was exposed at work exactly a week ago -- we've had an outbreak that has sent us remote. I teach -- many of my students and some coworkers have tested positive. I live with a family member that I would prefer not to expose to the virus. I've been quarantined for the last week and I've tested at home, and was negative 5 days (around what was the mean day of symptom onset) after exposure. I've had COVID before, and I got sick around 5 days after COVID exposure the last time. I'm triple vaccinated now, too. I still feel like I'm just waiting for the shoe to drop. Every twinge in my back, sneeze, and throat tickle puts me on edge (I'm in New England so the air is insanely dry right now, both outside due to the bitter cold and inside due to heating). I know I'm not out of the possible incubation period, but I just feel like I can't be so lucky as to have direct exposure and not get sick this time. Don't get me wrong, I'm very grateful for the vaccine, but it feels like with so many breakthrough cases that I should just expect to get reinfected. I've read booster efficacy sits around 75%. I feel like that should be more reassuring than what my current state of mind has me feeling. I used to feel a sense of comfort, having been vaccinated and having "hybrid" immunity, etc., but this surge of Omicron has really crushed that feeling. I know it's milder, which I'm happy about. It's just that this moment we're living through is incredibly unnerving. I'm essentially just venting and trying not to catastrophize. Thanks for reading.

Annotator 1

- Anticipation (Intensity: Anticipation)

→ Extractive: I know I'm not out of the possible incubation period, but I just feel like I can't be so lucky as to have direct exposure and not get sick this time.
→ Abstractive: I expect that I definitely have COVID for a second time because I was directly exposed during an outbreak at work and I can't be lucky enough to not get sick.

- Fear (Intensity: Fear)

→ Extractive: Every twinge in my back, sneeze, and throat tickle puts me on edge (I'm in New England so the air is insanely dry right now, both outside due to the bitter cold and inside due to heating).
→ Abstractive: I'm so afraid that I have COVID for a second time that every single one-off COVID-like symptom I have is putting me on edge.

- Sadness (Intensity: Sadness)

→ Extractive: I used to feel a sense of comfort, having been vaccinated and having "hybrid" immunity, etc., but this surge of Omicron has really crushed that feeling.
→ Abstractive: I'm sad that I don't have any more comfort from being vaccinated because Omicron has totally wiped that out.

Annotator 2

- Anticipation (Intensity: Anticipation)

→ Extractive: I'm very grateful for the vaccine, but it feels like with so many breakthrough cases that I should just expect to get reinfected.
→ Abstractive: I expect that I can still catch COVID even though I am fully vaccinated.

- Fear (Intensity: Fear)

→ Extractive: I still feel like I'm just waiting for the shoe to drop. Every twinge in my back, sneeze, and throat tickle puts me on edge.
→ Abstractive: I am afraid that these minor issues could be COVID symptoms, knowing that there is still a chance that I could have it.

- Sadness (Intensity: Pensiveness)

→ Extractive: I've read booster efficacy sits around 75%. I feel like that should be more reassuring than what my current state of mind has me feeling.
→ Abstractive: I can't help but feel a bit blue knowing there is a chance, no matter how small, that I can still get COVID.

Figure C.11: Example of the dataset. The bracketed levels indicate the emotion intensities.