7-2024

Unveiling the dynamics of crisis events: Sentiment and emotion analysis via multi-task learning with attention mechanism and subject-based intent prediction

Phyo Yi WIN MYINT
Singapore Management University, ywmphyo@smu.edu.sg

Siaw Ling LO
Singapore Management University, sllo@smu.edu.sg

Yuhao ZHANG
Singapore Management University, yuhaozhang@smu.edu.sg

Information Processing and Management 61 (2024) 103695

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Unveiling the dynamics of crisis events: Sentiment and emotion analysis via multi-task learning with attention mechanism and subject-based intent prediction

Phyo Yi Win Myint, Siaw Ling Lo *, Yuhao Zhang

Singapore Management University, 80 Stamford Road, 178902, Singapore

ARTICLE INFO

Keywords:
Multi-task learning
Natural language processing
Sentiment analysis
Emotion analysis
Intent prediction
Crisis tweets

ABSTRACT

In the age of rapid internet expansion, social media platforms like Twitter have become crucial for sharing information, expressing emotions, and revealing intentions during crisis situations. They offer crisis responders a means to assess public sentiment, attitudes, intentions, and emotional shifts by monitoring crisis-related tweets. To enhance sentiment and emotion classification, we adopt a transformer-based multi-task learning (MTL) approach with attention mechanism, enabling simultaneous handling of both tasks, and capitalizing on task interdependencies. Incorporating attention mechanism allows the model to concentrate on important words that strongly convey sentiment and emotion. We compare three baseline models, and our findings show that BERTweet outperforms the standard BERT model and exhibits similar performance to RoBERTa in crisis tweets. Furthermore, we employ natural language processing techniques to extract key subject entities (e.g., police, victims) and leverage the publicly available commonsense knowledge model, COMET-ATOMIC 2020, to identify their intentions in given crisis scenarios. Evaluation of COMET-ATOMIC 2020 on subject-based intent prediction in crisis tweets reveals that BART was superior to GPT2-XL model, providing crisis responders with vital information for better decision making. Notably, the integration of sentiment and emotion classification, identification of attention words and subject-based intent prediction represents a novel methodology, not previously applied in the context of crisis scenarios.

1. Introduction

In recent years, social media platforms have gained importance as key sources of information and communication during crisis events, including natural disasters, terrorist attacks, and public health emergencies. People often use platforms like Twitter to share experiences, express emotions and share critical information. The abundance of user-generated content on these platforms has created new opportunities for analyzing public sentiment and emotional responses during crises.

The analysis of sentiment and emotions in crisis-related tweets has garnered significant attention in natural language processing and computational linguistics (Beigi et al., 2016; Kaur & Kumar, 2015; Torkildson et al., 2014). Sentiment analysis assesses the overall polarity (positive, negative, neutral) of tweets to understand the overall attitude, while emotion analysis categorizes specific emotions. Understanding sentiment and emotional dynamics during crises is vital for various stakeholders, including emergency response

* Corresponding author.
E-mail address: sllo@smu.edu.sg (S.L. Lo).

https://doi.org/10.1016/j.ipm.2024.103695
Received 15 August 2023; Received in revised form 16 February 2024; Accepted 20 February 2024
Available online 8 March 2024

organizations, government agencies, and humanitarian aid providers (which we term as 'crisis responders' in this paper). Monitoring sentiment and emotions in crisis tweets helps crisis responders detect sudden shifts in sentiment, including increases or decreases in negative emotions. These insights aid in evaluating public perceptions and attitudes toward the crisis, making informed decisions, prioritizing crisis response efforts, communicating crisis management strategies, and effectively allocating resources.

Social media conversations not only convey sentiment and emotion but also various intents. One key factor of crisis management is understanding the intentions and needs of key entities, allowing crisis responders to provide targeted support and assistance. Subject-based intent prediction identifies crisis-related key entities (e.g., police force, victims, or assailants, which we termed as 'subject' in this paper) and predicts their intents. By focusing on these subjects and intents, crisis responders can gain deeper insights into the diverse range of needs, concerns, and goals of key entities during the crisis.

Due to the aforementioned motivation, we emphasize the importance of sentiment and emotion analysis and subject-based intent prediction in crisis situations. We adopt a Multi-task Learning (MTL) approach with an attention mechanism for sentiment and emotion classification. We train the model to handle two tasks simultaneously, leveraging the shared representations and task interdependencies. Furthermore, we employ Natural Language Processing (NLP) techniques to identify key subject entities and leverage a commonsense knowledge model, COMET-ATOMIC 2020, for subject-based intent prediction. We focus on tweets related to urban crisis events (e.g., civil disorder, armed assault) and aim to provide crisis responders (e.g., police and paramedics) with critical information for more effective crisis management and response strategies. The main contributions of this study are:

1. We compare the performances of three transformer-based MTL models, BERT, RoBERTa and BERTweet with attention mechanism for sentiment and emotion classification in crisis-related tweets.
2. We employ NLP techniques and leverage COMET-ATOMIC 2020, a commonsense knowledge model to determine the intent of key entities in urban crisis situations. To the best of our knowledge, our proposed approach was the first to unveil the dynamics of crisis events from a multi-facet perspective.

The rest of the paper is organized as follows: Section 2 discusses relevant literature, focusing on MTL with attention mechanism and subject-based intent prediction. Section 3 presents the proposed methodology. Section 4 explains the experiment setup, while Section 5 evaluates the experimental results. In Section 6, a case study and key insights from the approach are discussed. Section 7 addresses challenges and future work, and Section 8 provides the conclusion.

2. Related work

2.1. Crisis informatics and analysis

Crisis informatics studies the role of information and communication technology in responding to crises (Palen & Anderson, 2016). The use of social media during a crisis is becoming an emerging research topic in crisis informatics (Eilander et al., 2016; Imran et al., 2020; Shah & Schweiggart, 2023). Individuals actively utilize social media during crises to share vital information such as early warnings, safety precautions, and updates on infrastructure damage which are crucial for crisis management and communication efforts (Imran et al., 2020). Moreover, studies indicate that higher usage of social media is linked to more effective crisis management by government authorities (Graham et al., 2015). In recent years, extensive research in crisis informatics has made notable advancements in various areas such as crisis event detection, situational awareness, geoparsing, damage assessment and sentiment and emotional analysis (Imran et al., 2020). Poblete et al. (2018) proposed an online method to detect unusual spikes in discrete-time signals extracted from Twitter to detect earthquakes. Transfer learning technique with augmented data was utilized to detect wildfires using image data (Sousa et al., 2020). Agarwal et al. (2020) introduced a multi-modal sequential damage identification and severity detection system using the impact of linguistic cues on unimodal visual system to support crisis management. Jin and Spence (2023) analyzed communication trends during different stages of Hurricane Maria using topic modeling and latent semantic analysis. Karimzadeh et al. (2019) developed a geoparser using Named Entity Recognition (NER) tools for toponym recognition and a GeoNames gazetteer (a geographical dictionary) for toponym resolution. These contributions deepen our understanding of the effectiveness of leveraging social media data in crisis informatics, demonstrating their applicability and relevance across various types of crises.

Big social media data mining provides timely insights into public's opinion, views and attitudes regarding crisis events which could potentially impact crisis management strategies (Shah & Schweiggart, 2023). In today's world, social media has the capability to change public perceptions and their interpretation of crisis. Combining text mining, topic modeling and sentiment analysis, Shah and Schweiggart (2023) analyzed the temporal and topical shifts in public discussion to extract insights from data corresponding to each of the disaster based on the extended Fink's four-stages of crisis and disaster model. Rahman et al. (2023) employed text mining and sentiment analysis to identify ongoing social crises using Twitter data. Dai et al. (2024) created a sentiment dictionary for crises to identify the netizen emotion and used Latent Dirichlet Allocation (LDA) model for topic mining. Their research analyzed China Eastern Airlines MU5735 crash through the division of public opinion cycles and employing topic classification, sentiment-topic dynamic collaborative analysis and creating a visual public opinion map (Dai et al., 2024). Sentiment and emotion analysis was also used to analyze public opinion in the context of public security during crisis (Suhaimin et al., 2023). Bashir et al. (2021) conducted an analysis of tweet content and sentiments during and after Khan Shaykhun Syrian Chemical Attack by visualizing sentiments and creating network-based maps featuring keywords and word frequency for enhancing situational awareness and providing virtual social and emotion support. de Las Heras-Pedrosa et al. (2020) investigated sentiment and emotion analysis within Spanish society during the

COVID-19 pandemic to understand how social media influences crisis communication and its impact on digital ecosystems. Overall, these studies highlight the significance of utilizing social media for sentiment and emotion analysis in crisis informatics and their potential to aid government and crisis responders in crisis management. This has influenced and motivated our study, prompting us to explore how we can effectively integrate social media data, sentiment, emotion and subjected-based intent analysis into our own research to contribute to the advancements of crisis informatics and enhance crisis management.

2.2. MTL and attention mechanism in sentiment and emotion analysis

MTL has been proven effective in machine learning domains, including natural language processing(Collobert & Weston, 2008), speech recognition (Deng et al., 2013) and computer vision (Girshick, 2015). It often employs hard parameter sharing, utilizing a shared module across all tasks, followed by task-specific modules (Caruana, 1997). Hard parameter sharing is particularly advantageous for mitigating overfitting, making it valuable for tasks with limited training data (Caruana, 1997), which is well-suited in our context since labelled urban crisis data is not abundant.

In the crisis domain, Wang et al. (2020) introduced a multi-modal multi-task model using a single multi-modal dataset containing labels for different tasks. MTL was employed to identify location mentions in crisis tweets (Khanal & Caragea, 2021). Kumari et al. (2021) introduced a deep MTL framework which simultaneously handled novelty detection, emotion recognition and misinformation detection, demonstrating the superiority of MTL over single-task learning. Similarly, Akhtar et al. (2019) and Chauhan et al. (2020) presented a MTL model for sentiment and emotion analysis for multi-modal inputs. Experimental results from Akhtar et al. (2019) and Chauhan et al. (2020) showed the effectiveness of MTL in simultaneously modeling sentiment and emotion analysis, even for multi-modal inputs. The experimental findings from these studies (Akhtar et al., 2019; Chauhan et al., 2020; Kumari et al., 2021) consistently indicated that MTL model outperforms single-task models. Hence, this study adopts MTL to classify sentiments and emotions, aiming to gain a better understanding of crisis event dynamics.

In recent years, following the introduction of the attention-based transformer model by Vaswani et al. (2017), several transformer-based variations, including BERT, ELECTRA and T5, have emerged, achieving state-of-the-art performance in numerous language tasks through transfer learning (Kotei & Thirunavukarasu, 2023). Attention mechanism has been widely adopted in deep learning models, allowing them to selectively focus on different parts of input data during both training and inference stages (Kotei & Thirunavukarasu, 2023; Vaswani et al., 2017). It is used in neural machine translation, allowing the model to attend to relevant parts of the input sequence to increase accuracy (Choi et al., 2018). Furthermore, Liu et al. (2022) applied attention mechanism to identify witnesses of real-world events, utilizing tweets from crime investigation, rumor dispelling and emergency management. Attention mechanism has since been integrated into various MTL frameworks to improve task performance and model interpretability (Choi et al., 2018; Kotei & Thirunavukarasu, 2023; Vaswani et al., 2017). MTL with attention mechanism also showed promise in domains such as semantic segmentation (Zhao et al., 2021) and computer vision (Guo et al., 2022) achieving improved performance by addressing multiple tasks simultaneously. In the healthcare domain, MTL framework with attention mechanism was utilized to predict multiple disease outcomes from electronic health records, demonstrating superior performance compared to single-task models (Amjad et al., 2023). To the best of our knowledge, MTL with attention mechanism has not been previously applied to interpret sentiment and emotion in evolving crisis events. In this study, we extend the application of MTL to sentiment and emotion classification, specifically focusing on crisis-related tweets. We leverage attention mechanism to identify important words contributing to classification decisions, ultimately enhancing our understanding and interpretability of the model's predictions.

2.3. Subject-based intent analysis

Intent is defined as a purposeful action (Malle & Knobe, 1997; Sloman et al., 2012). We associate intent with various daily behaviors, including conversational interactions and information sharing. Several factors influence how individuals express intent, such as beliefs, desires, perceptions, and awareness (Malle & Knobe, 1997; Sloman et al., 2012). Earlier studies focused on extracting transactional intent related to commercial interests (Carlos & Yalamanchi, 2012; Hollerit et al., 2013; Wang et al., 2015). These limited motives were primarily associated with intentions related to transactions, such as buying and selling. However, these transactional intents differ from the intentions relevant to our current focus on urban crisis events. Intentions in urban crisis events, such as seeking help or protecting affected individuals, are more complex and require a more comprehensive examination. Most studies on social media data have concentrated on binary intent classification, both in commercial and crisis domains, due to the challenges of predicting intent in noisy text (Purohit et al., 2015). Consequently, multiclass, and open intent classification remain ongoing and challenging problems (Zhang et al., 2022). In the crisis domain, previous studies addressed issues such as problem-aid report (Varga et al., 2013), and request-offer messages but relied solely on binary classifiers (Purohit et al., 2013, 2014). Considering the complexity nature of urban crisis events, a classifier may struggle to capture the diverse intents and granularity related to key entities. Hence, in our study, we explore the use of the publicly available COMET-ATOMIC 2020 knowledge model which employs common-sense representation and reasoning to identify intent from key entities in crisis tweets.

3. Methodology

3.1. Workflow

Fig. 1 provides an overview of our proposed approach. The Crisis Detection System identifies tweets relevant to urban crisis events, such as "fire/explosion," "flood/typhoon," "civil disorder," "armed assault," and "bombing", from the noisy content. Specifically, we implement a crisis detection system based on BERTweet-LSTM-CRF (Zhang et al., 2023). Since our study focuses on downstream processing of crisis related tweets, we do not delve into the inner workings of the crisis detection system.

Crisis tweets are filtered and analyzed together at a predefined time interval, e.g., every 15 min. Each time interval is a point (or bin) in the sentiment/emotion analysis vs time. Therefore, the sentiments/emotions carried by the crisis tweets in a window represent the overall sentiments/emotions in that respective window.

We employ a MTL model with attention mechanism for sentiment and emotion analysis. By using attention mechanism built within transformers, we can extract words with highest attention (weights) and use these attention words to provide insights during prediction. The outputs of the MTL model include sentiment and emotion classifications with their respective attention words.

Subject detection extracts key subject entities (e.g., police, victims) from the crisis tweets. We then group similar subjects using cosine similarity and assign them to specific clusters. To leverage the COMET-ATOMIC 2020 model, crisis related tweets are preprocessed as required by the model to generate the intent associated with the subject entity. The resulting outputs include subject entities, intent, and the cluster numbers to which the subjects are assigned.

Overall, our approach provides a comprehensive workflow which integrates crisis detection, time-based tweet selection, sentiment, and emotion MTL classification with attention mechanism, subject detection, and subject cluster assignments with intent prediction.

3.2. Multi-task learning

3.2.1. Model architecture

Our study employs inductive transfer learning by leveraging a pre-trained language model which is fine-tuned on a diverse range of Twitter data (BERTweet) (Nguyen et al., 2020). We utilize a simultaneous learning approach, where different tasks are learned concurrently using a shared representation. Training on multiple tasks helps the model learn a more general representation, leading to improved performance across tasks (Choi et al., 2018; Kotei & Thirunavukarasu, 2023; Vaswani et al., 2017). Fig. 2 illustrates the specific architecture employed for MTL in our study.

The input sequence, consisting of a tweet, is tokenized into individual tokens. Position embeddings, crucial for representing the position of words in a token sequence, are added to each token. These embeddings allow the model to distinguish between tokens based on their position in the text and capture the semantic meaning and contextual information of individual tokens. Token embeddings, along with position embeddings, are processed through a shared encoder. The encoder processes the token and position embeddings, generating contextual embeddings for each token. These embeddings capture contextual information and token relationships. The shared encoder, consisting of 12 layers of self-attention and feed-forward neural networks, generates a contextualized representation, denoted as “o”, and represents the token’s meaning in the context of the entire input text. The pooled output is subsequently generated by applying max pooling to the “o” embeddings. Max pooling selects the highest activation value from the “o” representation of each token, which are aggregated to create the pooled output. The pooled output is then passed through a task-specific layer which is a fully connected layer for task-specific predictions. The sentiment layer produces predictions for positive, negative, or neutral sentiment while the emotion layer generates predictions for emotion, such as love, joy, sadness, anger, surprise, fear, or no emotion.

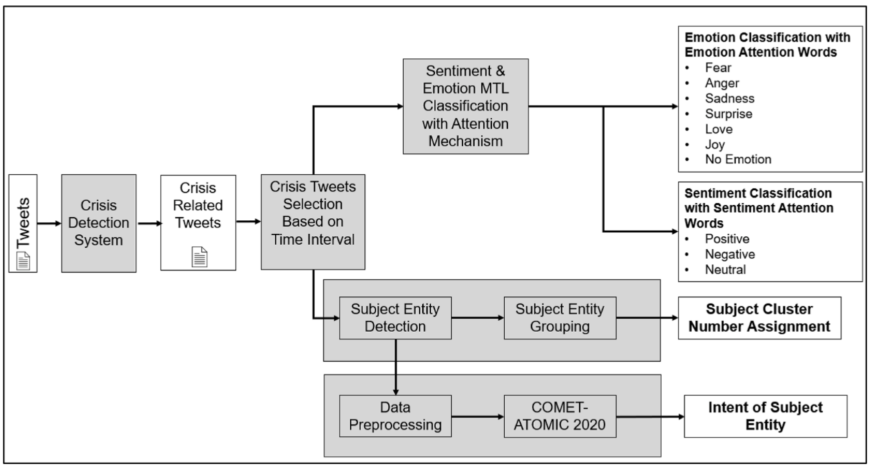

Fig. 1. Workflow diagram of the proposed approach.

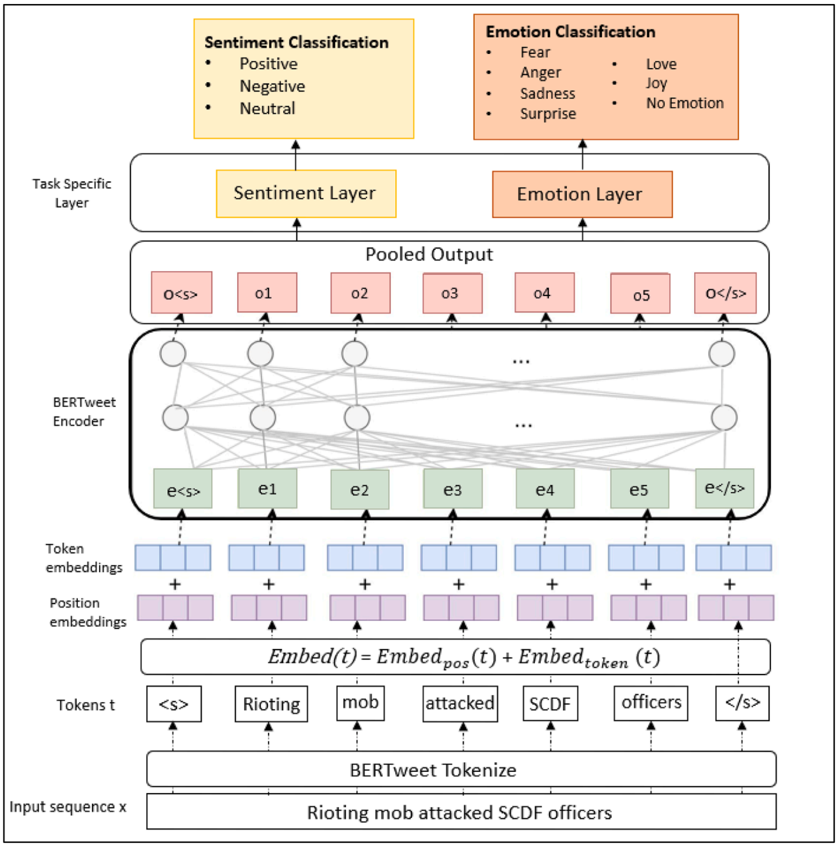

Fig. 2. Overview architecture of multi-task learning approach.

3.2.2. Loss function of MTL

In MTL, a single model is trained to handle multiple tasks simultaneously which can sometimes lead to competition between tasks. To address this issue, we establish a loss function that effectively trains the MTL model while preventing one task from dominating the other. Different tasks require distinct loss functions; for instance, classification tasks use cross-entropy, while regression tasks employ mean squared error.

In our approach, we use the weighted sum of each loss function in MTL, with the weights determined by their homoscedastic uncertainty (Kendall et al., 2018). Homoscedastic uncertainty, also known as task-specific uncertainty, allows each task to have its own estimated uncertainty. It means that MTL model acknowledges that some tasks might inherently produce noisier or more uncertain predictions than others. By incorporating task-specific uncertainty into the MTL framework, the model can balance the contributions of each task to the overall training objective. Tasks with higher uncertainty are assigned lower weight in the optimization process, while tasks with lower uncertainty are given higher weights. This balancing approach ensures that no single task dominates the training process. It also helps manage interdependencies between two tasks without conflicts and has been proven effective, especially in tasks involving scene geometry and semantics (Kendall et al., 2018). While their research primarily focuses on regression tasks, we have adapted their method for a classification problem.

Specifically, we use cross-entropy loss to calculate the loss for both sentiment and emotion classification. Additionally, we incorporate precision (inverse of the variance) into the loss formula, with the logarithm of the variance parameter representing the uncertainty for each task. The total loss function is computed as the sum of the precision-weighted losses.

The cross-entropy loss for Task 1 (L_1) can be written as:

$$L_1 = \sum_i y_i \log(p_i) \textit{for all samples } i \textit{ in Task } 1$$

The cross-entropy loss for Task 2 (L_2) can be written as:

$$L_2 = \sum_j z_j \log(q_j) \textit{for all samples } j \textit{ in Task } 2$$

The precision-weighted loss for Task 1 (WL_1) can be written as:

$$WL_1 = precision_1 \ . \ L1 + \log(var_1)$$

The precision-weighted loss for Task 2 (WL_2) can be written as:

$$WL_2 = precision_2 \ . \ L2 + \log(var_2)$$

The total multi-task loss for both tasks can be defined as the sum of the precision-weighted losses:

$$TL = WL_1 + WL_2$$

Task 1 and Task 2 correspond to sentiment and emotion classification, respectively. y_i and p_i are the ground truth label and predicted probability of the correct class for the i th sample in Task 1, respectively. z_j and q_j are the ground truth label and predicted probability of the correct class for the j-th sample in Task 2, respectively. $precision_1$ and $precision_2$ are the precision value for Task 1 and Task 2, respectively. var_1 and var_2 are the variance value associated with precision for Task 1 and 2, respectively.

3.2.3. Attention mechanism

The attention mechanism focuses on specific input segments during both training and inference stages (Vaswani et al., 2017). In our approach, we employ an attention mechanism based on the BERTweet Transformer architecture, which utilizes self-attention using three vectors: query, key, and value. The vectors are derived from the input word embeddings (Luo et al., 2020; Vaswani et al., 2017).

To calculate attention scores between words, we compute a dot product between the query vector of a word and the key vectors of all the words in the sequence. The dot product measures the similarity between the query and the key vectors, resulting in a vector of attention scores. These scores are then normalized into probabilities using a Softmax function, serving as weights for the value vectors. Each value vector is multiplied by its corresponding attention weight and these weighted value vectors are summed up. The result is a weighted sum of the value vectors, allowing the model to focus on an important span of the input tweet for each task.

After predicting sentiment and emotion, we leverage the attention mechanism to extract words closely related to the predictions by aggregating the weights from all 12 attention heads of the transformer. We use the 75th percentile value as a threshold to determine word importance, achieving a balance between information retention and filtering out less relevant words. We then sort the words weights in descending order, from the highest to the lowest. Words with higher weights indicate greater importance and analyzing these weights provides insights into individual word importance and their impact on the model's sentiment and emotion predictions.

3.3. Subject-based intent prediction

3.3.1. COMET-ATOMIC 2020

The COMET-ATOMIC 2020 model is a language generation model which is trained using the COMmonsensE Transformer (COMET) framework and fine-tuned on the ATOMIC dataset, a large-scale commonsense repository of textual descriptions (Hwang et al., 2021). The COMET framework provides a comprehensive approach to train language models in generating sentences that incorporate commonsense knowledge (Hwang et al., 2021).

The COMET-ATOMIC 2020 model is designed to capture various types of relations, offering a total of 23 relation types such as xNeed, xEffect, xReact, and xWant, among others (Hwang et al., 2021). However, our specific focus in this study is on the "xIntent" relation type to analyze the intentions of the subject.

3.3.2. Subject extraction

In crisis tweets, we extract subjects (e.g., crisis victims or suspects) using Part-of-Speech (POS) tagging and dependency parsing. These techniques analyze the grammatical structure of tweets before identifying the subject. POS tagging is used to label each word in a tweet with its corresponding part-of-speech category, such as noun, verb, adjective, etc. Dependency parsing analyzes the syntactic relationships between words in a tweet and creates a dependency tree which represents how words depend on each other for their roles within the tweet. Our method identifies the word or phrase that fulfills the role of the subject based on its syntactic position and POS tag. Common dependency labels associated with subjects are used to extract the subject entities in the tweets. Examples are as follows.

1. "nsubj" (nominal subject):

   "`Bus driver knocked down the Indian`." (Subject: "Bus driver")

2. "nsubjpass" (passive nominal subject):

   "`The riot was started by Indian workers`." (Subject: "Indian workers")

3. "csubj" (clausal subject):

   "`What the person is injured is uncertain`" (Subject: "the person")

4. "csubjpass" (passive clausal subject):

   "`Whether emergency responders arrive timely is yet to be determined.`"(Subject: "emergency responders")

3.3.3. Tweet formating

We observe that the format of the training data for COMET-ATOMIC 2020 always starts with Person X, e.g. “Person X washes the car”, “Person X spills the tea”. Hence, we modify the format of the tweets in order to leverage the COMET-ATOMIC 2020 model more effectively.

If a subject is extracted, as discussed in Section 3.3.2, it is replaced with “Person X”. If no subject is detected in the tweet, a prefix “Person X” is added at the start of the tweet. The following two examples explain the additional formating:

Original Tweet 1: `Man held over gun attack on police`.

Formated Tweet 1: Person X held over gun attack on police.

In Tweet 1, a subject “Man” is detected and thus “Man” is replaced with “Person X”.

Original Tweet 2: `stab the man to death`.

Formated Tweet 2: Person X stab the man to death.

In Tweet 2, there is no subject detected. In this case, a prefix “Person X” is added in front of the tweet.

3.3.4. Subject entity grouping

If subjects such as “policeman” and “police” which are synonyms or have similar meanings, are not properly managed, they can be treated as separate subjects due to expression variations. To address this, we group subjects together and assign them cluster numbers using the algorithm in Fig. 3. We use “en_core_web_md” (English Core Web Medium) which contains pre-trained word vectors, to calculate cosine similarity between pairs of subjects. Subjects with cosine similarity above a defined threshold are grouped together and assigned the same cluster number. A higher cosine similarity threshold indicates greater similarity between subjects. However, there are instances where a subject is not assigned to any cluster. This is due to the absence of a subject entity in the tweet, or the subject entity does not belong to any existing cluster. Since there are not enough representatives of the subject in the cluster, the subject is likely not a key entity in the event and thus omitted from further analysis.

```
Algorithm: Subject Entity Grouping and Cluster Assignment
Input: subject entities S, cosine similarity threshold k
Output: groups of similar subject entities based on cosine similarity G, Cluster T
G, T = { } , { }
// Step 1: Get word embedding of each subject
E(s) = embedding for each subject s in subject entities S
// Step 2: Check if subject is already in a group
for each subject s in subject entities S:
    for each group g in groups G:
        if subject s in group g:
            in_a_group = True
            break
    // Step 3: if the subject s is not in a group, create a new group for it
    if not in_a_group:
        g' = [s]
        // Go through other remaining subjects and add them to the current group if similar
        for other subject s' in subject entities S:
            if s' = s:
                continue
            // Step 4: Check if the other subject s in already in a group
            for each group g in groups G:
                if each subject s' in group g:
                    in_a_group = True
                    break
            // Step 5: If the other subject is not in a group, retrieve its embedding and
            compute its similarity to the current subject
            if not in_a_group:
                similarity = cosine-similarity(E(s), E(s'))
                // if the similarity is above the threshold, add the subject to the current group
                if similarity > threshold k:
                    g'. append(s')
        // Step 6: add the new group to the list of groups
        G. append(g')
// Step 7: Assort the group g in groups G based on the size of g in descending order
G = sorted(g)
// Step 8: Assign cluster number
for each group g in G:
    Assign g to cluster index T
Return G, T
```

Fig. 3. Subject entity grouping and cluster assignment algorithm.

Table 1
Emotion annotation rule.

Emotion	Sense of Emotion Tone	Example of Emotion Related Words
Fear	Apprehension, Unease, Concern	Terrified, Scared, Anxious, Worried
Anger	Frustration, Annoyance, Hostility	Frustrated, Irritated, Annoyed
Sadness	Sorrow, Grief, Melancholy	Depressed, Mournful, Grieving
Surprise	Amazement, Astonishment, Shock	Shocked, Amazed, Surprised
Love	Affection, Adoration, Tenderness	Cherish, Devoted, Adored
Joy	Happiness, Elation, Delight	Happy, Excited, Delighted
No Emotion	No emotion conveyed	None

3.4. Annotation

3.4.1. Sentiment and emotion annotation

Sentiment is classified into three categories, Negative, Positive and Neutral, based on the overall tone and mood of the tweet (Shaver et al., 1987). A tweet with a cheerful, happy, enthusiastic, or excited tone is labeled as "Positive" (Shaver et al., 1987). A tweet conveying anger, annoyance or frustration is labeled as "Negative" (Shaver et al., 1987). "Neutral" sentiment is assigned to tweets which do not convey any positive or negative emotions, attitudes, or opinions (Shaver et al., 1987). For example, "he went to the mall yesterday" is assigned as "Neutral" sentiment.

We adopt the Shaver Emotion Model (Table 1) which is based on six emotions such as fear, anger, sadness, surprise, love and joy (Shaver et al., 1987). Additionally, we add "No Emotion" category for tweets which contain factual information without emotional or evaluative content. As explained above, "he went to the mall yesterday" is annotated as "Neutral" and "No Emotion".

3.4.2. Intent annotation

For evaluation of COMET-ATOMIC 2020 model, we manually annotate the intention of 200 crisis and non-crisis related tweets based on our common sense and contextual knowledge. The following tweet example shows the annotated and predicted intent of the subject entity.

Tweet: `A bus hit an Indian guy.`

Predicted Intent: [to hurt someone, to be violent, to be mean]

Annotated Intent: [to hurt an Indian guy]

4. Experiment

4.1. Setup for multi-task learning

4.1.1. Dataset

The dataset consists of Sentiment Dataset (Sentiment140, n.d.), Emotion Dataset (Saravia et al., 2018), Crisis Related Dataset (Olteanu et al., 2015) and Local Crisis Dataset, extracted via Twitter API as shown in Table 2. Sentiment140 provides sentiment labels related to brands, products, or topics from Twitter, making it valuable for understanding public sentiment towards specific entities or subjects. Emotion dataset includes English tweets with six basic emotions covering a broad range of topics and events. CrisisLexT26 comprises tweets related to global crisis events such as typhoons, wildfires, and earthquakes. Since this study focuses on urban crisis events and aims to analyze local crisis events, we include two local Singapore events and extract their relevant dataset for the study. Little India Riot dataset is a large-scale, complex riot incident involving multiple event categories such as armed assault, fire, and traffic accident, while Orchard Slashing dataset is a small-scale armed assault incident. We apply the annotation guidelines outlined in Section 3.4.1 to label Crisis Related Dataset (CrisisLexT26) and Local Crisis Dataset. After combining all the datasets from different sources shown in Table 2, Tables 3 and 4 show the final breakdown of sentiment and emotion dataset used for the MTL model.

4.1.2. Data pre-processing for MTL

We tokenize tweets by splitting them into individual words while removing punctuation marks like hashtags. Subsequently, we proceed with stop word removal, eliminating common words such as "and" or "the" which do not carry significant sentiment and emotion information. Numbers are also removed to focus on textual content. We normalize all mentions of username and URL, replacing "@user" with the token "@USER" and hyperlink with the token "HTTPURL". Additionally, we replace emojis with their aliases. For instance, the emoji 😂 is replaced by the token ":face_with_tears_joy:" using emoji Python library. Lastly, we replace multiple consecutive spaces with single spaces and replace line breaks with a space. We retain capital letters as they are sometimes used to emphasize certain words in tweets, indicating stronger sentiment or emotion.

4.1.3. Model configuration

We finetune the MTL model using three baseline models with the same transformer-based architecture: BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (A Robustly Optimized BERT Pretraining Approach) and BERTweet (BERT based model trained on Twitter data).

Table 2
Twitter datasets used for MTL of sentiment and emotion.

Data	Label Type	No. of Tweets		Data Source
Sentiment Dataset	Positive	512	1391	Sentiment 140
	Negative	356		
	Neutral	523		
Emotion Dataset	Fear	431	3363	Hugging Face
	Anger	496		
	Sadness	485		
	Surprise	502		
	Love	511		
	Joy	487		
	No Emotion	451		
Crisis Related Dataset	Typhoon	31	245	CrisisLexT26
	Wildfires	35		
	Earthquake	37		
	Floods	38		
	Bushfire	35		
	Shooting	36		
	Bombing	33		
Local Crisis Dataset	Little India Riot	478	954	Twitter API
	Orchard Slashing	476		
Total			5953	

Table 3
Sentiment dataset for multi-task learning model.

Sentiment	No. of Tweets	% of Crisis Tweets	% of non-Crisis Tweets
Positive	1804	11 %	89 %
Negative	2651	46 %	54 %
Neutral	1498	15 %	85 %

Table 4
Emotion dataset for multi-task learning model.

Emotion	No. of Tweets	% of Crisis Tweets	% of non-Crisis Tweets
Fear	835	34 %	66 %
Anger	945	29 %	71 %
Sadness	781	31 %	69 %
Surprise	720	23 %	77 %
Love	639	10 %	90 %
Joy	677	12 %	88 %
No Emotion	1356	24 %	76 %

BERT model (“bert-base-cased) is the most commonly used transformed-based model, pretrained on a corpus containing 3.3 billion words. RoBERTa (“roberta-base”) is an improved version of BERT which employs an optimized training process, utilizing larger batch sizes and dataset compared to BERT. As a result, it achieves state-of-the-art performance in various NLP tasks making it a suitable choice as one of the baseline models for comparison. BERTweet is a specialized variation of BERT trained on Twitter data to better understand the informal language and unique characteristics of tweets. This makes it a good fit for our experiment, given that our source data consists of tweets.

We split the data into training, testing and validation data with a ratio of 70:20:10, respectively. We train on training split data and evaluate on validation split data after each training epoch. At the end of training, it is evaluated on testing split data. The maximum token length was set to 128. The loss function is as described in Section 3.2.2. The batch size and learning rate are set to 16 and 2e-5, respectively. The models are optimized using Adam optimizer and trained for 10 epochs.

4.2. Setup of COMET-ATOMIC 2020

4.2.1. Dataset and data pre-processing for COMET-ATOMIC 2020

The testing dataset consists of 100 crisis tweets from CrisisLexT26 dataset and 100 non-crisis tweets obtained through Twitter API. We apply the annotation guideline in Section 3.4.2 to label the testing dataset. We apply the same data processing steps to the tweets as we do in the MTL approach described in Section 4.1.2. Subsequently, we extract subjects using the method in Section 3.3.2 and format the tweets as specified in Section 3.3.3.

4.2.2. Model setup

We perform experiments using two publicly available models: COMET-ATOMIC 2020 BART and GPT2-XL. Both models leverage knowledge from ATOMIC 2020 dataset, with one adopting BART (Bidirectional and Auto Regressive Transformers) while the other integrates the capabilities of GPT2 (Generative Pre-Trained Transformer 2).

5. Results and discussion

5.1. Evaluation of MTL

5.1.1. Evaluation metrics

Typical accuracy metrics used in binary classification focus on true positive and true negative and have known limitations in terms of reflecting the performance of a classifier (Sokolova et al., 2006). Accuracy can be misleading when dealing with imbalanced datasets, which is found in our scenario. Hence, we choose F1 and Macro-F1 metrics over accuracy since they account for both precision and recall, providing a more comprehensive evaluation of the classifier's performance.

The F1 score is the harmonic mean of both precision and recall, where precision is defined as the ratio of true positives found from the predicted positives, while recall is the ratio of true positives identified from the actual positives (Sokolova et al., 2006).

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

The Macro F1 is the average F1 score across all classes, providing comprehensive insights into the model's overall effectiveness.

$$Macro - F1 = \frac{Sum\ of\ F1\ Scores\ for\ all\ classes}{number\ of\ classes}$$

5.1.2. Sentiment and emotion prediction results

We compare three transformer-based MTL models: RobBERTa, BERT and BERTweet, as shown in Tables 5–7. To maintain consistency, we preprocess tweets for all models as described in Section 4.1.2.

Among the three models, BERTweet consistently achieves the highest F1 score across all sentiment classes. BERT and RoBERTa perform similarly in both "Negative" and "Neutral" sentiment categories, achieving F1 scores of 0.92 and 0.82, respectively. In "Positive" sentiment class, BERT and BERTweet perform similarly with F1 score of 0.91 while RoBERTa follows closely with F1 score of 0.90.

For emotion classification, RoBERTa obtains the best F1 score for "Anger", "Fear", "Sadness" and "Surprise", while BERTweet achieves the highest F1 score in predicting "Joy" and "Love". Both BERTweet and RoBERTa perform similarly in "No Emotion" category with F1 score of 0.89. We observe neither BERTweet nor RoBERTa achieve the worst performance of F1 score for any class of emotions whereas BERT has comparatively lower performance in all seven emotion classes.

In sentiment classification, BERTweet outperforms BERT and RoBERTa, achieving the best Macro-F1 score of 0.90. In emotion classification, BERTweet and RoBERTa perform similarly, obtaining the highest Macro-F1 score of 0.86.

Table 8 summarizes the time to complete the training (10 epochs) and evaluation on the testing split data. All models run in the same computing environment (Nvidia GeForce RTX 3070, 16GB RAM). All three models have relatively similar training times with BERTweet being the fastest (10 min and 23 s). In addition, BERTweet stands out as the quickest to evaluate, taking only 46 s.

Since BERTweet achieves the best Macro-F1 score for both sentiment and emotion classification and computationally more efficient, we incorporate BERTweet as the final model in our subsequent study.

5.1.3. Analysis of sentiment and emotion classification

It is important to note that emotions can be context dependent and expressed differently by different individuals in various situations. There are instances where some tweets carry primary emotions and secondary emotions. However, since our MTL model focuses on predicting a single emotion, it is possible the model sometimes predicts a secondary emotion instead of a primary emotion even though the sentiment remains the same, i.e., the primary emotion can be "Fear" and the secondary can be "Anger", but both are associated with "Negative" sentiment. This could explain why the performance of emotion classification is lower than sentiment classification. The following two tweets are examples of how tweets can carry both primary and secondary emotions:

Tweet 1: `Riot at Little India. Police car was smashed earlier. Please avoid that area and give way to emergency vehicles!`

In Tweet 1, the annotated emotion is "Anger", but the predicted emotion is "Fear".

Tweet 2: `Omg I hope nobody gets too hurt in the riot at Little India:(`

In Tweet 2, the annotated emotion is "Fear", but the predicted emotion is "Sadness".

Fig. 4 presents an overview of the distribution of emotions within each sentiment category. Notably, the sentiment category

Table 5
Sentiment prediction performance on testing split dataset.

Model	Negative			Neutral			Positive		
	P	R	F1	P	R	F1	P	R	F1
BERT	0.92	0.92	0.92	0.78	0.85	0.82	0.94	0.87	**0.91**
BERTweet	0.96	0.91	**0.93**	0.74	0.93	**0.83**	0.98	0.84	**0.91**
RoBERTa	0.96	0.88	0.92	0.72	0.94	0.82	0.98	0.83	0.90

Table 6
Emotion prediction performance on testing split dataset.

Model	Emotion	P	R	F1
BERT	Anger	0.63	0.80	0.70
BERTweet		0.83	0.80	0.81
RoBERTa		0.80	0.86	**0.84**
BERT	Fear	0.79	0.73	0.76
BERTweet		0.81	0.84	0.83
RoBERTa		0.84	0.83	**0.84**
BERT	Sadness	0.82	0.83	0.83
BERTweet		0.82	0.90	0.86
RoBERTa		0.88	0.89	**0.89**
BERT	Joy	0.70	0.86	0.77
BERTweet		0.78	0.93	**0.84**
RoBERTa		0.73	0.91	0.81
BERT	Love	0.93	0.83	0.88
BERTweet		0.94	0.93	**0.94**
RoBERTa		0.94	0.84	0.89
BERT	Surprise	0.89	0.82	0.85
BERTweet		0.85	0.88	0.86
RoBERTa		0.88	0.87	**0.88**
BERT	No Emotion	0.90	0.76	0.82
BERTweet		0.97	0.83	**0.89**
RoBERTa		0.95	0.84	**0.89**

Table 7
Comparison of MTL models on testing split dataset.

Model	Macro-F1	
	Sentiment	Emotion
BERT	0.89	0.80
BERTweet	**0.90**	**0.86**
RoBERTa	0.89	**0.86**

Table 8
Comparison of training and evaluation time for MTL models.

Model	Training Time	Evaluation Time
BERT	00:10:53	00:00:56
BERTweet	**00:10:23**	**00:00:46**
RoBERTa	00:10:45	00:01:02

"Negative" is primarily associated with the emotions of "Anger" and "Fear," which have the highest percentages at 32.35 % and 25.00 %, respectively. “Sadness” follows closely with a percentage of 16.18 %. Conversely, “Negative” sentiment category has lower percentages of the emotions “Love” and “Joy”. In contrast, the "Positive" sentiment category is strongly associated with positive emotions, particularly "Joy" and "Love", which make up 31.65 % and 28.48 %, respectively. The emotions "Fear," "Anger," and "Sadness" have relatively lower percentages within the "Positive" sentiment category. The "Neutral" sentiment category exhibits a notable predominance of the "No Emotion" category in terms of percentage distribution. Overall, the observed data suggests a significant correlation between sentiment and emotion.

5.1.4. Analysis of attention words

As part of our analysis, we investigate the attention words and their roles in influencing the model’s prediction. Our analysis is performed on testing split data which allows us to evaluate the generalizability of the model to new and unseen data. Table 9 shows the

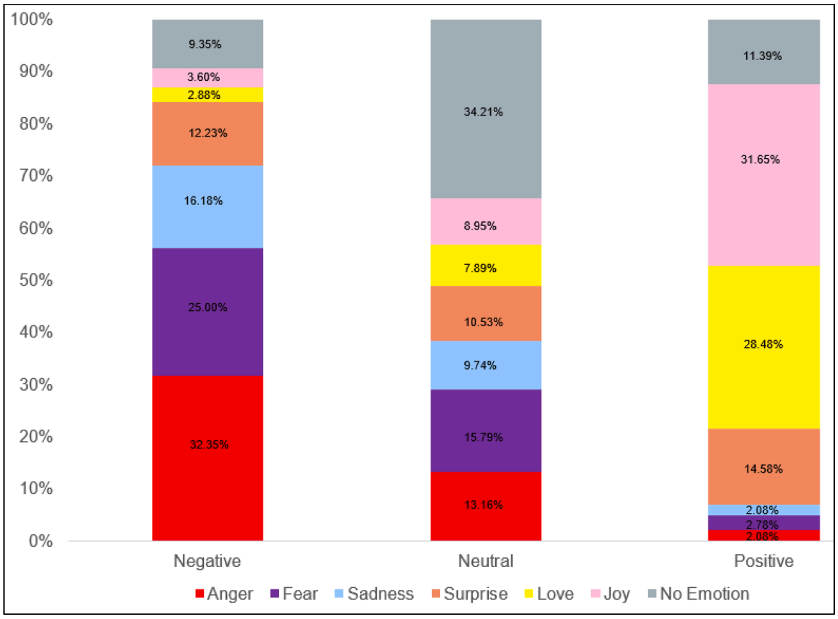

Fig. 4. Percentage breakdown of emotion by sentiment.

attention weights from the task-specific layer for randomly selected input tweets. In the visualization, words with darker colors indicate higher attention weights and greater significance in influencing the model's prediction.

We observe that the model exhibits the capability to effectively capture sentiment and emotion related words. In the first example, the model assigns the highest attention weights to the words "stranded" and "lost" in sentiment classification and these words truly convey a negative sentiment. For emotion classification, the model relies heavily on the words "stranded" and "lost" to predict the emotion of "Fear". These words are emotional indicators signifying the emotional response of fear. In the second example, the attention of the model is on "Haiyan", "Typhoon", "faces" when predicting "Neutral" and "No Emotion" class. These words do not carry any sentiment or emotional state but simply provide factual information about the typhoon. The model's selection of these words shows its ability to recognize the absence of sentiment and emotion content in the tweet. In the third example, the model identifies "respect", "firemen", "manage" in predicting "Positive" and "Love". These words convey a sense of appreciation and acknowledgement which also suggests a "Positive" sentiment. This shows the ability of the model to detect positive and joyful emotions by recognizing words associated with admiration and appreciation.

5.2. Evaluation of intent prediction

5.2.1. Evaluation metrics

We employ three commonly used evaluation metrics for text generation: BLEU score (Papineni et al., 2002), METEOR score (Lavie & Denkowski, 2009), and BERT score (Zhang et al., 2019). The same evaluation metrics are also used in the original COMET-ATOMIC 2020 model when it is first introduced (Hwang et al., 2021). All three metrics evaluate machine-generated text by comparing it to a reference text. In our context, machine-generated text is the predicted intent from COMET-ATOMIC 2020 model and reference text is the annotated intent. The metric value ranges from 0 to 1, where a higher value means a closer match between the predicted and annotated intent.

BLEU (Bilingual Evaluation Understudy) employs precision as a key metric and includes a brevity penalty to account for text length differences between the annotated intent and the predicted intent. Precision measures the degree of overlap of n-grams (where n is the maximum n-gram size considered, e.g., 1 for unigrams, 2 for bigrams, etc.) between the predicted and annotated intent. In our context, we use unigrams ($n = 1$). BLEU score formula is as follows:

$$BLEU = Brevity\ Penalty\ \times \left(\prod_{i=1}^{n} Precision_i\right)^{\frac{1}{n}}$$

METEOR (Metric for Evaluation of Translation with Explicit ORdering) score considers aspects such as stemming, synonyms and word order (Banerjee & Lavie, 2005; Lavie & Denkowski, 2009). First, unigram precision (P) measures the ratio of the number of unigrams in the predicted intent that are mapped (to unigrams in the annotated intent) to the total number of unigrams in the predicted intent. Similarly, unigram recall (R) is computed as the ratio of the number of unigrams in the predicted intent that are mapped (to unigrams in the annotated intent) to the total number of unigrams in the annotated intent. Next, we compute F_{mean} by combining precision and recall using a harmonic mean, with more weight given to recall. Unigrams mapped to annotated unigrams are grouped

Table 9
Visualization of attention words from BERTweet MTL model on testing split data.

	Tweet	Prediction
1	The rescue team is searching for people lost in the devastating flood. They are also trying to rescue people stranded by rising waters.	Negative
	The rescue team is searching for people lost in the devastating flood. They are also trying to rescue people stranded by rising waters.	Fear
2	1 month after Super Typhoon Haiyan Philippines faces huge challenges.	Neutral
	1 month after Super Typhoon Haiyan Philippines faces huge challenges.	No Emotion
3	So much respect for the firemen, police officers and all those who are there to manage the situation.	Positive
	So much respect for the firemen, police officers and all those who are there to manage the situation.	Love

into chunks. The penalty increases as the number of chunks increases. Penalty is calculated based on the number of chunks, reducing F_{mean} maximum of 50 % if there are no bigram or longer matches. Precision, recall and F_{mean} are based on unigram matches. The METEOR score formula is as follows:

$$F_{mean} = \frac{10PR}{R + 9P}$$

$$Penalty = 0.5 \times \left(\frac{\#chunks}{\#unigrams\_matched}\right)^3$$

$$Score = F_{mean} \times (1 - Penalty)$$

BERT (Bidirectional Encoder Representations from Transformers) score leverages the pre-trained contextual embeddings from BERT (Zhang et al., 2019) and converts the predicted and annotated intent into embedding using BERT. It performs tokenization and computes the cosine similarity between the embeddings of the predicted and annotated intent and uses F1 score which combines precision and recall assessing the quality of token-level overlap between the predicted and annotated intent.

For each tweet, we compute the similarity between the predicted intent and the annotated intent and then select the best matching score pair. For all 200 tweets, we calculate the average scores of BLEU, METEOR and BERT as shown in Table 10.

5.2.2. Intent prediction results

The findings from Table 10 indicate that COMET-ATOMIC 2020 BART outperforms COMET-ATOMIC 2020 GPT2-XL in intent prediction. We observe the evaluation time of BART (36 s) is faster than GPT2-XL (2 min 14 s). Hence, we choose COMET-ATOMIC 2020 BART as the final model in our subsequent study.

5.2.3. Analysis of intent prediction

While attention words contribute to the contextual understanding of tweets and their classification, we also observe instances where the intent prediction effectively clarifies the situation, going beyond the role of attention words as shown in Table 11.

In the first example from Table 11, the attention words capture the challenging situation of the earthquake. However, the predicted intent offers a more comprehensive understanding of the rescue teams' primary goals, such as saving lives, providing hope, and offering support during the aftermath of earthquake. Similarly, the second example of predicted intent provides a more complete interpretation of the aftermath of the wildfire compared to the attention words alone.

6. Case study evaluation: Singapore Little India Riot 2013

6.1. Background

In this section, we present a case study of Singapore Little India Riot to explain the effectiveness of our proposed approach. Singapore Little India Riot occurred on December 8, 2013 at 21:23 local time, in the Little India district of Singapore. The riot was triggered by a private bus accident resulting in the death of an Indian construction worker. A crowd quickly gathered at the accident scene, and the situation turned violent as the rioters attacked the private bus and emergency vehicles. Authorities responded by deploying police reinforcements and Singapore Civil Defense Force (SCDF) officers to disperse the mob and arrest rioters. The riot involved approximately 300 migrant workers and resulted in significant damage to 25 emergency vehicles, with five of them set on fire.[1]

6.2. Understanding Little India Riot through the proposed approach

As shown in Table 12, rioters and SCDF officers are the two main significant clusters detected in the riot. Rioters are central figures in instigating and fueling the riot, whereas SCDF officers provide emergency response including evaluating the situation, providing medical attention, extinguishing fires, and assisting those in need. Analyzing the sentiment and emotion expressed in tweets which mention either rioters or SCDF officers provides better understanding of the riot dynamics. To do this, we specifically select tweets which mention only one type of subject.

6.3. Uncovering insights from Rioter related tweets

Firstly, we gather tweets from December 8, 21:00 to December 9, 3:00. These tweets are processed through a crisis detection system to identify and filter tweets related to Little India Riot. We group tweets into 30-minute intervals which allow us to track trends and patterns of the riot. Our analysis, presented in Table 12, Figs. 5 and 6 shows that the predominant sentiment conveyed regarding the riot and the rioters is "Negative" while the primary emotion is "Anger". Notably, during the initial stages of the riot, the number of "Negative" sentiment and "Anger" emotion tweets are relatively low. However, as the crisis escalates, the sentiment and emotion of the

[1] https://en.wikipedia.org/wiki/2013_Little_India_riot

Table 10
Evaluation results of intent prediction.

Model	BLEU	METEOR	BERT Score
COMET-ATOMIC 2020 (BART)	**0.40**	**0.32**	**0.66**
COMET-ATOMIC 2020 (GPT2-XL)	0.33	0.24	0.61

Table 11
Output of the proposed approach.

Tweet	Sentiment, Sentiment Attention Words	Emotion, Emotion Attention Words	Subject	Intent
Rescue teams working tirelessly to find survivors after earthquake… in this challenging time.	Neutral [tirelessly, challenging, working, survivors]	Fear [challenging, survivors, tirelessly, earthquake]	Rescue teams	[to save lives, to help people, to be helpful]
Communities are coming together to share resources and help evacuate those in danger after wildfire crisis.	Neutral [evacuate, coming, danger, share]	Fear [evacuate, danger, wildfire, share, coming]	communities	[to help others, to help people]

Table 12
Output of the proposed approach from 21:00 to 22:00.

Tweet	Sentiment, Sentiment Attention Words	Emotion, Emotion Attention Words	Subject	Intent	Cluster
Rioting mob attacked SCDF rescuers	Negative [rioting, rescuers, attacked]	Anger [rioting, rescuers, attacked]	Rioting mob	[to get revenge, to hurt people, to be cause trouble]	1
Singapore Little India Riot 2013 - Mob Surrounds Vehicles #singapore	Negative [surrounds, Mob, little, riot]	Anger [surrounds, Mob, riot, vehicles]	Mob	[to express anger, to cause trouble]	1
Angry mob overturns Police cars & sets them on fire. Updated with vids. Riot in Little India	Negative [overturns, angry, sets]	Anger [updated, riot, angry, overturns]	Angry mob	[to express anger, to get revenge]	1
SCDF officers are really brave. Going out to the riot and they can even lose their lives	Neutral [brave, SCDF, lose]	Surprise [riot, lose, brave, lives]	SCDF officers	[to be brave, to protect the people]	2
SCDF officers standby and shields up. Racecourse road condone off. Apparently, a bus hit an Indian guy	Neutral [Racecourse, hit, standby, condone]	Fear [condone, standby, Indian]	SCDF officers	[to be safe, to protect people, to help]	2
SCDF officers personnels controlling the streets of Little India tonight, you've got a big job on your hands. Stay strong and stay safe, mates	Positive [personnels, safe, controlling]	Joy [controlling, safe, stay]	SCDF officers personnels	[to protect people, to protect the city, to be helpful]	2
Bus driver ran over a local. No response from emergency services for over 30mins. Bus driver attacked.	Negative [attacked, ran, driver, response]	Anger [attacked, ran, emergency]	Bus driver	[to get revenge, to hurt someone]	3
China bus driver knocked down an Indian guy and he died.	Negative [died, knocked, driver]	Fear [knocked, died, Indian]	China Bus driver	[to kill someone, to hurt someone]	3
oh, apparently a bus driver hit down a Bangla. Then riot ah	Negative [hit, riot, Bangla]	Anger [Bangla, hit, riot]	Bus driver	[to cause trouble, to hurt someone]	3

rioters intensify, reaching their peak between 23:00 and 00:30. As seen from Table 12, tweets under cluster "1″ are associated with angry mob and their intentions primarily revolve around protesting, hurting people and getting revenge for the migrant worker who was run over by a bus driver. From Table 12, Figs. 5 and 6, crisis responders can anticipate the presence of violent and angry rioters, enabling them to develop effective strategies to manage such individuals. The attention words such as "rioting", "attacked", "rescuers", "overturns", "vehicles" and "angry" emphasize the violent nature of the riot and align with the intent of the rioters. This highlights the underlying motives of the rioting mob which include seeking retribution and causing harm to people.

6.4. Uncovering insights from SCDF officers related tweets

Similarly, we conduct sentiment and emotion analysis regarding SCDF officers during the period from December 8, 21:00 to December 9, 3:00, using 30-min intervals as shown in Figs. 7 and 8. In contrast to Fig. 6, which predominantly shows "Anger" and "Fear" emotions, Fig. 8 shows a diverse range of emotions in tweets related to SCDF officers. While "Anger" and "Fear" persist throughout the riot, other emotions such "Sadness", "Love", "Joy" and "Surprise" are also present. Examining these tweets and their attention words such as "injured", "worried", "sorry", "damage" reveals that mentions of injured SCDF officers during the riot evokes

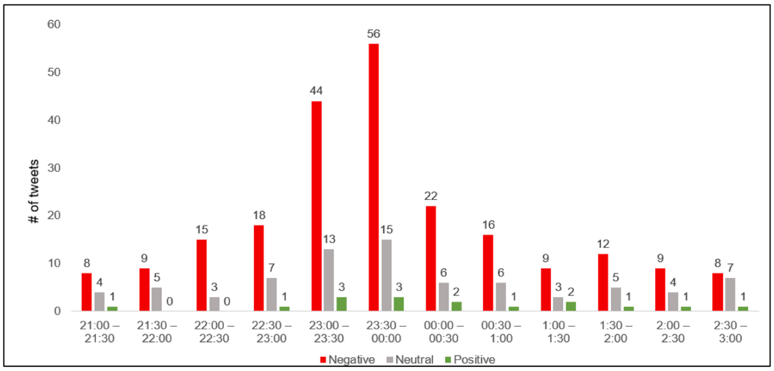

Fig. 5. Sentiment analysis of tweets mentioning rioters from 21:00 to 3:00.

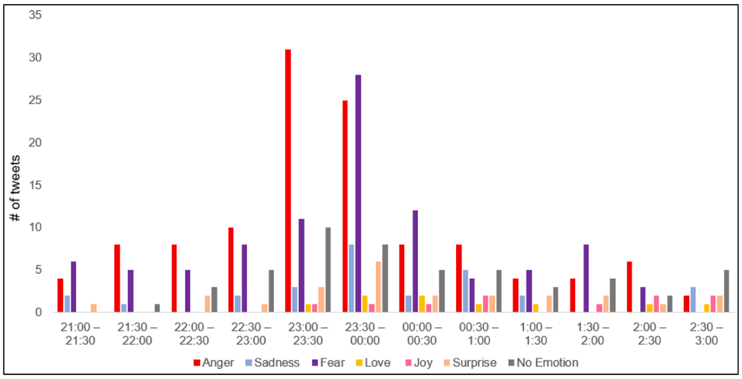

Fig. 6. Emotion analysis of tweets mentioning rioters from 21:00 to 3:00.

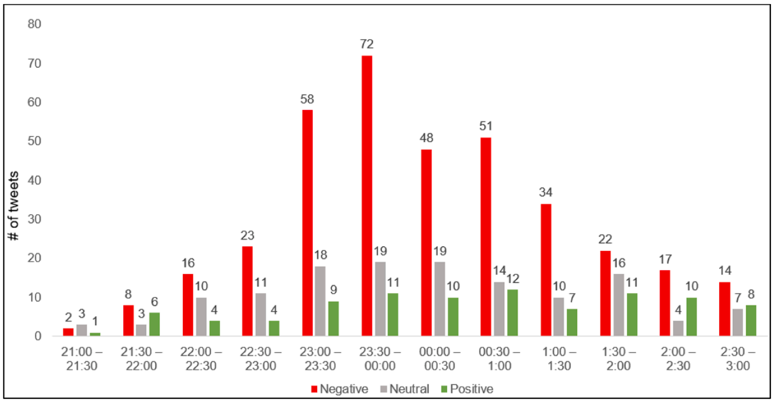

Fig. 7. Sentiment analysis of tweets mentioning SCDF officers from 21:00 to 3:00.

feelings of sadness among citizens. Additionally, tweets with attention words such as "strong," "brave," "safe," and "justice" in relation to SCDF officers express love and joy, indicating citizens' pride and appreciation for their actions. This suggests that citizens are content with how SCDF officers manage the riot. The attention words "condone", "standby", "brave", "safe", "controlling" in Table 12 highlight the preparedness and safety measures taken by SCDF officers. The intents of SCDF officers such as "to be safe", "to protect people", "to protect the city" and "to be helpful" also align with theses attention words, emphasizing SCDF's role in maintaining order and safety.

6.5. Actionable recommendations for crisis responders

Crisis management involves prevention, planning, response, recovery, and learning (Prayag, 2018). Crisis responders' ability to respond effectively is highly dependent on their ability to obtain actionable information about the on-the-ground situation during the

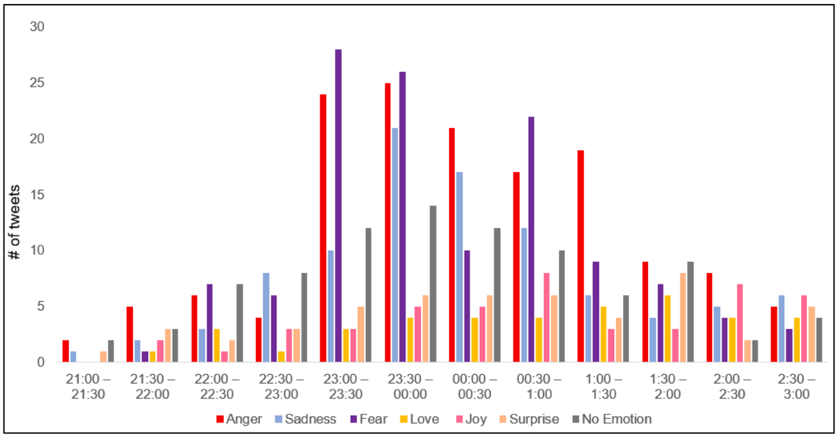

Fig. 8. Emotion analysis of tweets mentioning SCDF officers from 21:00 to 3:00.

response stage (McCreadie et al., 2020). Therefore, monitoring and understanding public response, sentiment and emotion during a crisis is crucial. Not only does it aid in understanding the issues individuals are facing, but it also allows for an evaluation of the government's response to the crisis (Imran et al., 2020). Overall, understanding the sentiment and emotion of the key subject entities provides real-time insights into the emotional state of entities affected by the riot. Pluwak et al. (2023) analyzed the sentiment and emotion of public opinions to detect possible warning signs of crisis in near real-time. Similarly, the extracted sentiment and emotion help to enable responders to monitor and detect early signs of distress from the public in this study. Consequently, this equips crisis responders with better situational awareness for making informed and timely decisions during the response stage (Pluwak et al., 2023). Additionally, recognizing key entities in the riot such as victims, protestors, and authority, helps crisis responders understand the complex dynamics of the riot. By understanding the intentions of the key entities, crisis responders can provide targeted response, assistance, and support such as medical and emotional aid and logistical support to the specific key entities during response and recovery stage (Imran et al., 2020). Effective crisis communication is essential during both response and recovery phases as it shapes public attitudes towards crisis events (Xie et al., 2022). Through social media, emergency information such as evacuations, warnings, and reassurances can be disseminated (Xie et al., 2021). Thus, crisis communication not only impacts the development of crisis events and the effectiveness of crisis management, but also has the potential to influence public perceptions, emotions, and attitudes towards crisis (Xie et al., 2021). Having access to timely information regarding public sentiment is crucial for government agencies and crisis responders to formulate effective strategies for managing the situation (Kydros et al., 2021). This knowledge also aids in crafting precise communication guidelines on social media to distribute accurate and reliable information, thereby ensuring public safety (Kydros et al., 2021). Research has shown that intentions (Xie et al., 2021), sentiment and emotion analysis (Beigi et al., 2016) can be valuable tools in post-crisis evaluation. Hence, the data gathered and analyzed by our proposed approach can be used for post-crisis analysis and evaluation, which aids in identifying areas of improvement in crisis management strategies and communication approaches leading to better preparedness for future crises.

7. Challenges and future work

7.1. Social media biases

Since our study relies on social media data, it is important to consider that there are biases related to the type of the users and the language they use. For biases related to the type of users, we intend to explore two primary avenues in our future work. First, we plan to conduct a detailed demographic analysis of the user profiles and study how these demographics may impact the sentiment, emotion and intent expressed in tweets. Secondly, we aim to develop user-based sentiment and emotion models that can account for these demographic variations.

One of the primary challenges in sentiment and emotion analysis is capturing the contextual nuances of the language. Textual content is highly context-dependent, and accurately interpreting sentiment and emotion requires an understanding of the specific context in which the text is generated. Multilingual and cross-cultural perspectives must be considered since different languages and cultures express sentiments and emotions uniquely. Future research should focus on building multilingual and cross-cultural sentiment and emotion analysis frameworks by incorporating local knowledgebase, leveraging techniques such as transfer learning, multilingual embeddings, and domain adaptation.

Sarcasm and irony pose significant challenges in sentiment and emotion analysis since they involve words or phrases that convey the opposite of their literal meaning. Future work should explore novel approaches, such as incorporating discourse analysis, pragmatic reasoning, or advanced machine learning techniques, to better handle the detection and interpretation of sarcasm and irony.

Tweets contain noisy, informal text with abbreviations, slang, misspellings, and disregard for proper grammar rules. This presents challenges not only for sentiment and emotion analysis but also for extracting subject entities from tweets. Future work should explore techniques to efficiently preprocess and normalize noisy and informal text to enhance the performance of sentiment and emotion analysis model and improve subject entity extraction in tweets.

7.2. Attention mechanism

Our analysis of the attention mechanism emphasizes the significance of attention words in driving the model's predictions. As explained in Section 5.2.3, predicted intents provide a more comprehensive depiction of the crisis scenario compared to the attention words. Hence, future research can focus on investigating attention bias in our model and refining the attention mechanism to capture important words more accurately through the implementation of adaptive attention weights, multi-head attention and incorporating contextual information to improve the model's sensitivity.

7.3. COMET-ATOMIC 2020

COMET-ATOMIC 2020 BART is trained on general cause and effect scenarios with specific structure (for example, assuming the subject or 'X' to always precede a sentence – X votes for Y, X runs out of steam etc.) (Hwang et al., 2021). In this study, we have pre-processed the data in order to fit into the structure for a better prediction. Future work should involve incorporating crisis related tweets as training data by retraining COMET-ATOMIC 2020 BART model to work well without the need to pre-process the data.

8. Conclusion

In conclusion, we propose a transformer-based multi-task learning approach with attention mechanism to enhance the performance of sentiment and emotion classification in crisis tweets, by leveraging shared representations and interdependencies between tasks. Three models, namely BERT, BERTweet and RoBERTa are assessed in classifying sentiment and emotion from the crisis tweets with BERTweet, achieving the highest Macro-F1. We incorporate advanced natural language processing techniques such as POS tagging, and dependency parsing to extract subject entities and employ COMET-ATOMIC 2020 BART for subject-based intent prediction. By carrying out sentiment and emotion analysis, valuable insights can be obtained regarding on-the-ground situations, aiding in the planning of crisis management strategies and crisis response efforts. Furthermore, subject-based intent prediction provides a deeper understanding of the diverse needs and goals of key entities involved in the crisis, facilitating better decision making and the implementation of targeted response strategies by crisis responders.

CRediT authorship contribution statement

Phyo Yi Win Myint: Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. **Siaw Ling Lo:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing. **Yuhao Zhang:** Data curation, Methodology, Validation, Writing – review & editing.

Data availability

Data will be made available upon request but the authors do not have permission to share the source codes.

Acknowledgements

The authors would like to acknowledge the support and project funding from ST Engineering Mission Software & Services Pte Ltd under Research Collaboration Agreement No: 001052-00001.

References

Agarwal, M., Leekha, M., Sawhney, R., & Shah, R. R. (2020). Crisis-DIAS: Towards multimodal damage analysis - deployment, challenges and assessment. In *, 34*. *Proceedings of the AAAI conference on artificial intelligence* (pp. 346–353). https://doi.org/10.1609/aaai.v34i01.5369

Akhtar, M. S., Chauhan, D., Ghosal, D., Poria, S., Ekbal, A., & Bhattacharyya, P. (2019). Multi-task learning for multi-modal emotion recognition and sentiment analysis. In *Proceedings of the conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 370–379). https://doi.org/10.18653/v1/N19-1034

Amjad, H., Ashraf, M., Sherazi, S., Khan, S., Fraz, M., Hameed, T., et al. (2023). Attention-based explainability approaches in healthcare natural language processing. In *Proceedings of the 16th international joint conference on biomedical engineering systems and technologies*. https://doi.org/10.5220/0011927300003414

Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL workshop on intrinsic and extrisic evaluation measures for machine translation and/or summarization* (pp. 65–72).

Bashir, S., Bano, S., Shueb, S., Gul, S., Mir, A. A., Ashraf, R., et al. (2021). Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria chemical attack. *International Journal of Disaster Risk Reduction, 62*, Article 102397. https://doi.org/10.1016/j.ijdrr.2021.102397

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. *Sentiment Analysis and Ontology Engineering*, 313–340. https://doi.org/10.1007/978-3-319-30319-2_13

Carlos, C. S., & Yalamanchi, M. (2012). Intention analysis for sales, marketing, and customer service. In *Proceedings of the computational linguistics international conference* (pp. 33–40).

Caruana, R. (1997). Multitask learning. *Machine Learning, 28*(1), 41–75. https://doi.org/10.1023/a:1007379606734

Chauhan, D. S., S R, D., Ekbal, A., & Bhattacharyya, P. (2020). Sentiment and emotion help sarcasm? A multi-task learning framework for multi-modal sarcasm, sentiment and emotion analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 4351–4360). https://doi.org/10.18653/v1/2020.acl-main.401

Choi, H., Cho, K., & Bengio, Y. (2018). Fine-grained attention mechanism for neural machine translation. *Neurocomputing, 284*, 171–176. https://doi.org/10.1016/j.neucom.2018.01.007

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing. In *Proceedings of the 25th international conference on machine learning - ICML 2008* (pp. 160–167). https://doi.org/10.1145/1390156.1390177

Dai, J., Zhao, Y., & Li, Z. (2024). Sentiment-topic dynamic collaborative analysis-based public opinion mapping in aviation disaster management: A case study of the MU5735 air crash. *International Journal of Disaster Risk Reduction, 102*, Article 104268. https://doi.org/10.1016/j.ijdrr.2024.104268

de Las Heras-Pedrosa, C., Sánchez-Núñez, P., & Peláez, J. I. (2020). Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. *International Journal of Environmental Research and Public Health, 17*(15), 5542. https://doi.org/10.3390/ijerph17155542

Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. In *Proceedings of rhe IEEE international conference on acoustics, speech and signal processing* (pp. 8599–8603). https://doi.org/10.1109/icassp.2013.6639344

Eilander, D., Trambauer, P., Wagemaker, J., & Van Loenen, A. (2016). Harvesting social media for generation of near real-time flood maps. *Procedia Engineering, 154*, 176–183. https://doi.org/10.1016/j.proeng.2016.07.441

Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1440–1448). https://doi.org/10.1109/iccv.2015.169

Graham, M. W., Avery, E. J., & Park, S. (2015). The role of social media in local government crisis communications. *Public Relations Review, 41*(3), 386–394. https://doi.org/10.1016/j.pubrev.2015.02.001

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., et al. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media, 8*(3), 331–368. https://doi.org/10.1007/s41095-022-0271-y

Hollerit, B., Kröll, M., & Strohmaier, M. (2013). Towards linking buyers and sellers: Detecting commercial intent on Twitter. In *Proceedings of the 22nd international conference on world wide web* (pp. 629–632). https://doi.org/10.1145/2487788.2488009

Hwang, J. D., Bhagavatula, C., Le Bras, R., Da, J., Sakaguchi, K., Bosselut, A., et al. (2021). (Comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In *, 35*. *Proceedings of the AAAI conference on artificial intelligence* (pp. 6384–6392). https://doi.org/10.1609/aaai.v35i7.16792

Imran, M., Ofli, F., Caragea, D., & Torralba, A. (2020). Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions. *Information Processing & Management, 57*(5), Article 102261. https://doi.org/10.1016/j.ipm.2020.102261

Jin, X., & Spence, P. R. (2023). Check crisis information on Twitter: Information flow and crisis communication patterns of hurricane ida. *Communication Studies, 74*(4), 337–355. https://doi.org/10.1080/10510974.2023.2202872

Karimzadeh, M., Pezanowski, S., MacEachren, A. M., & Wallgrün, J. O. (2019). GeoTxt: A scalable geoparsing system for unstructured text geolocation. *Transactions in GIS, 23*(1), 118–136. https://doi.org/10.1111/tgis.12510

Kaur, H., & Kumar, R. (2015). Sentiment analysis from social media in crisis situations. In *Proceedings of the international conference on computing, communication & automation* (pp. 251–256). https://doi.org/10.1109/CCAA.2015.7148383

Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7482–7491).

Khanal, S., & Caragea, D. (2021). Multi-task learning to enable location mention identification in the early hours of a crisis event. In *Proceedings of the findings of the association for computational linguistics: EMNLP* (pp. 4051–4056). https://doi.org/10.18653/v1/2021.findings-emnlp.340

Kotei, E., & Thirunavukarasu, R. (2023). A systematic review of transformer-based pre-trained language models through self-supervised learning. *Information, 14*(3), 187. https://doi.org/10.3390/info14030187

Kumari, R., Ashok, N., Ghosal, T., & Ekbal, A. (2021). Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition. *Information Processing & Management, 58*(5), Article 102631. https://doi.org/10.1016/j.ipm.2021.102631

Kydros, D., Argyropoulou, M., & Vrana, V. (2021). A content and sentiment analysis of Greek tweets during the pandemic. *Sustainability, 13*(11), 6150. https://doi.org/10.3390/su13116150

Lavie, A., & Denkowski, M. J. (2009). The METEOR metric for automatic evaluation of machine translation. *Machine Translation, 23*, 105–115. https://doi.org/10.1007/s10590-009-9059-4

Liu, W., Cao, J., Yang, Y., Liu, B., & Gao, Q. (2022). Category-universal witness discovery with attention mechanism in social network. *Information Processing & Management, 59*(4), Article 102947. https://doi.org/10.1016/j.ipm.2022.102947

Luo, X., Ding, H., Tang, M., Gandhi, P., Zhang, Z., & He, Z. (2020). Attention mechanism with BERT for content annotation and categorization of pregnancy-related questions on a community Q&A site. In *Proceedings of the IEEE international conference on bioinformatics and biomedicine (BIBM)* (pp. 1077–1081). https://doi.org/10.1109/bibm49941.2020.9313379

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*(2), 101–121. https://doi.org/10.1006/jesp.1996.1314

McCreadie, R., Cody Buntain, C., & Soboroff, I. (2020). Incident streams 2019: Actionable insights and how to find them. In *Proceedings of the 17th international conference on information systems for crisis response and management* (pp. 744–760).

Nguyen, D. Q., Vu, T., & Nguyen, A. T. (2020). BERTweet: A pre-trained language model for english tweets. In *Proceedings of the conference on empirical methods in natural language processing: system demonstrations* (pp. 9–14). https://doi.org/10.18653/v1/2020.emnlp-demos.2

Olteanu, A., Vieweg, S., & Castillo, C. (2015). What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing - CSCW '15* (pp. 994–1009). https://doi.org/10.1145/2675133.2675242 [Dataset].

Palen, L., & Anderson, K. M. (2016). Crisis informatics—new data for extraordinary times. *Science (New York, N.Y.), 353*(6296), 224–225. https://doi.org/10.1126/science.aag2579

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 311–318). https://doi.org/10.3115/1073083.1073135

Pluwak, A., Kacprzak, E., Stańczyk, M., Lew, M., & Obuchowski, A. (2023). Assessing the situation in the face of crisis. *Social Network Analysis and Mining, 13*(1), 151. https://doi.org/10.1007/s13278-023-01140-6

Poblete, B., Guzmán, J., Maldonado, J., & Tobar, F. (2018). Robust detection of extreme events using twitter: Worldwide earthquake monitoring. *IEEE Transactions on Multimedia, 20*(10), 2551–2561. https://doi.org/10.1109/TMM.2018.2855107

Prayag, G. (2018). Symbiotic relationship or not? Understanding resilience and crisis management in tourism. *Tourism Management Perspectives, 25*, 133–135. https://doi.org/10.1016/j.tmp.2017.11.012

Purohit, H., Castillo, C., Diaz, F., Sheth, A., & Meier, P. (2013). Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday, 19*(1). https://doi.org/10.5210/fm.v19i1.4848

Purohit, H., Dong, G., Shalin, V., Thirunarayan, K., & Sheth, A. (2015). Intent classification of short-text on social media. In *Proceedings of the IEEE international conference on smart city/SocialCom/SustainCom (SmartCity)* (pp. 222–228). https://doi.org/10.1109/SmartCity.2015.75

Purohit, H., Hampton, A., Bhatt, S., Shalin, V. L., Sheth, A. P., & Flach, J. M. (2014). Identifying seekers and suppliers in social media communities to support crisis coordination. *Computer Supported Cooperative Work (CSCW), 23*(4–6), 513–545. https://doi.org/10.1007/s10606-014-9209-y

Rahman, S., Jahan, N., Sadia, F., & Mahmud, I. (2023). Social crisis detection using Twitter based text mining-a machine learning approach. *Bulletin of Electrical Engineering and Informatics, 12*(2), 1069–1077. https://doi.org/10.11591/eei.v12i2.3957

Saravia, E., Liu, H. C. T., Huang, Y. H., Wu, J., & Chen, Y. S. (2018). CARER: Contextualized affect representations for emotion recognition. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 3687–3697). https://doi.org/10.18653/v1/D18-1404 [Dataset].

Sentiment140. (n.d.). Sentiment140- A Twitter Sentiment Analysis Tool. Retrieved from http://www.sentiment140.com/.

Shah, A. M., & Schweiggart, N. (2023). # BoycottMurree campaign on twitter: Monitoring public response to the negative destination events during a crisis. *International Journal of Disaster Risk Reduction, 92*, Article 103734. https://doi.org/10.1016/j.ijdrr.2023.103734

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*(6), 1061–1086. https://doi.org/10.1037//0022-3514.52.6.1061

Sloman, S. A., Fernbach, P. M., & Ewings, S. (2012). A causal model of intentionality judgment. *Mind & Language, 27*(2), 154–180. https://doi.org/10.1111/j.1468-0017.2012.01439.x

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *Proceedings of the Australasian joint conference on artificial intelligence* (pp. 1015–1021). https://doi.org/10.1007/11941439_114

Sousa, M. J., Moutinho, A., & Almeida, M. (2020). Wildfire detection using transfer learning on augmented datasets. *Expert Systems with Applications, 142*, Article 112975. https://doi.org/10.1016/j.eswa.2019.112975

Suhaimin, M. S. M., Hijazi, M. H. A., Moung, E. G., Nohuddin, P. N. E., Chua, S., & Coenen, F. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. *Journal of King Saud University-Computer and Information Sciences*. , Article 101776. https://doi.org/10.1016/j.jksuci.2023.101776

Torkildson, M. K., Starbird, K., & Aragon, C. (2014). Analysis and visualization of sentiment and emotion on crisis tweets. In *Proceedings of the 11th international conference on cooperative design, visualization, and engineering* (pp. 64–67). https://doi.org/10.1007/978-3-319-10831-5_9

Varga, I., Sano, M., Torisawa, K., Hashimoto, C., Ohtake, K., Kawai, T., et al. (2013). Aid is out there: Looking for help from tweets during a large scale disaster. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 1619–1629).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

Wang, J., Cong, G., Zhao, X., & Li, X. (2015). Mining user intents in Twitter: A semi-supervised approach to inferring intent categories for Tweets. In *, 29. Proceedings of the AAAI conference on artificial intelligence*. https://doi.org/10.1609/aaai.v29i1.9196

Wang, T., Tao, Y., Chen, S. C., & Shyu, M. L. (2020). Multi-task multimodal learning for disaster situation assessment. In *Proceedings of the IEEE conference on multimedia information processing and retrieval* (pp. 209–212). https://doi.org/10.1109/MIPR49039.2020.00050

Xie, C., Zhang, J., Huang, Q., Chen, Y., & Morrison, A. M. (2022). An analysis of user-generated crisis frames: Online public responses to a tourism crisis. *Tourism Management Perspectives, 41*, Article 100931. https://doi.org/10.1016/j.tmp.2021.100931

Xie, C., Zhang, J., Morrison, A. M., & Coca-Stefaniak, J. A. (2021). The effects of risk message frames on post-pandemic travel intentions: The moderation of empathy and perceived waiting time. *Current Issues in Tourism, 23*(23), 3387–3406. https://doi.org/10.1080/13683500.2021.1881052

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with BERT. In *Proceedings of the eighth international conference on learning representations 2020*.

Zhang, X., Cai, F., Hu, X., Zheng, J., & Chen, H. (2022). A contrastive learning-based task adaptation model for few-shot intent recognition. *Information Processing & Management, 59*(3), Article 102863. https://doi.org/10.1016/j.ipm.2021.102863

Zhang, Y., Lo, S. L., & Win Myint, P. Y. (2023). Impact of difficult noise on twitter crisis detection. In *Proceedings of the Pacific Asia conference on information systems*.

Zhao, Q., Liu, J., Li, Y., & Zhang, H. (2021). Semantic segmentation with attention mechanism for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–13. https://doi.org/10.1109/tgrs.2021.3085889