Graph-based Joint Pandemic Concern and Relation Extraction on Twitter

Jingli Shi[a], Weihua Li[a], Sira Yongchareon[a], Yi Yang[c], Quan Bai[b]

[a]*Engineering, Computer & Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*
[b]*Information & Communication Technology, University of Tasmania, Hobart, Australia*
[c]*School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China*

Abstract

Public concern detection provides potential guidance to the authorities for crisis management before or during a pandemic outbreak. Detecting people's concerns and attention from online social media platforms has been widely acknowledged as an effective approach to relieve public panic and prevent a social crisis. However, detecting concerns in time from massive volumes of information in social media turns out to be a big challenge, especially when sufficient manually labelled data is in the absence during public health emergencies, e.g., COVID-19. In this paper, we propose a novel end-to-end deep learning model to identify people's concerns and the corresponding relations based on Graph Convolutional Networks and Bi-directional Long Short Term Memory integrated with Concern Graphs. Except for the sequential features from BERT embeddings, the regional features of tweets can be extracted by the Concern Graph module, which not only benefits the concern detection but also enables our model to be high noise-tolerant. Thus, our model can address the issue of insufficient manually labelled data. We conduct extensive experiments to evaluate the proposed model by using both manually labelled tweets and automatically labelled tweets. The experimental results show that our model can outperform the state-of-the-art models on real-world datasets.

Keywords: Concern detection, COVID-19, Auto Concern Extraction, Concern Graph, Graph Convolutional Network

1. Introduction

The outbreak of coronavirus (COVID-19) in 2019 has been causing a rapid increase in both infection and death rates around the world. Especially when

Email addresses: `jingli.shi@aut.ac.nz` (Jingli Shi), `weihua.li@aut.ac.nz` (Weihua Li), `sira.yongchareon@aut.ac.nz` (Sira Yongchareon), `yyangjohnny@gmail.com` (Yi Yang), `quan.bai@utas.edu.au` (Quan Bai)

the pandemic moved into the second, third, or even fourth wave, it caused devastating loss of human life, impacted the global economy, transformed our daily lives, and posed a threat to our society (Killgore et al., 2020). According to the studies on the past pandemic outbreaks, e.g., Zika (Fu et al., 2016; Glowacki et al., 2016), Ebola (Lazard et al., 2015; Van Lent et al., 2017), and H1N1 (Chew & Eysenbach, 2010; Szomszor et al., 2011), social media platforms, e.g., Twitter, have proven to be a popular channel for spreading information, especially related to public opinions and concerns (Damiano & Catellier JR, 2020). This is because people tend to perceive more details regarding the pandemic by reading the newsfeeds and interpreting the comments from others through social networks (Li et al., 2018; Hu et al., 2019). Twitter, a popular and informative social network platform, allows people to post and interact with messages known as "tweets". They can also communicate and express opinions about the latest events (Killgore et al., 2020). User-generated tweets from Twitter turn out to be prophetic, namely, valuable indicators of what issues will likely happen in the pandemic. Therefore, it is important to make use of tweets and investigate what various people are discussing during the pandemic. The attitudes and behaviours of our society are affected directly by public concerns. Thus, how to effectively extract public concerns and analyse the corresponding relationships will assist people in understanding the anxiety and fears of the society in this pandemic situation. Furthermore, the potential social crisis can also be revealed by analysing public concerns, which significantly contribute to social management control.

Motivated by this background, great effort has been dedicated to mining social media data and exploring opinions towards pandemic outbreaks (da Silva et al., 2021). Most existing research works can be categorised into traditional survey methods, e.g., survey and questionnaire (Nelson et al., 2020), and machine learning model-based methods, e.g., topic modelling (Van Der Vegt & Kleinberg, 2020; Kassab et al., 2020). The existing studies are capable of extracting fundamental public concerns, e.g., "social distancing", "hand sanitiser" and "face masks", which require intensive human effort in labelling large datasets, turning out to be inefficient. Moreover, in any epidemic emergence situation, e.g., COVID-19, traditional approaches, such as questionnaires and clinical tests, neither collect enough data for deep learning model training nor rapidly generate a model for concern detection. Therefore, it is vital to design an end-to-end model that is capable of automatically analysing social media data and detecting public concerns without requiring a large-scale of data to be labelled manually.

Deep learning methods are increasingly applied to valuable information extraction. However, most methods rely heavily on data labelled by the annotators, requiring much time and financial resources (Kipf & Welling, 2017). Moreover, the noisy and imbalanced social media data prevent deep learning-based methods from generalisation (Rathan et al., 2018). In many existing studies, the proposed models are not able to track real-time statistics of public concerns related to pandemics due to the required labelled dataset (Li et al., 2020; Jahanbin et al., 2020; Hou et al., 2020; Lazard et al., 2015). To mitigate this issue, preliminary research was conducted to mine public concerns by proposing an Automated Concern Exploration (ACE) framework (Shi et al., 2021). The proposed framework can detect concerns from tweets automatically and construct a concern knowledge graph to present the interconnections of the extracted concern entity set. However, several advent limitations are still to be addressed. (1) only BERT embedding of tweets is used, which cannot capture regional dependency word features from tweets to improve the performance of concern extraction. (2) the relation between concerns in one tweet posted by a user is not detected, which is critical to reveal meaningful information about public concerns. (3) the framework employs a rule-based method, having poor generalisability and appearing difficult to transfer to future occurring pandemics.

In this paper, we propose and develop an end-to-end model with Concern Graph (CG) and concern states to simultaneously identify public concerns and corresponding relations. "Public concern" is formally defined with a consideration of its type and degree, and construct a concern graph to represent the regional features, improving the concern identification effectiveness. Furthermore, the proposed method can extract concern relations by integrating concern states with Graph Convolutional Network (GCN) (Kipf & Welling, 2017). Extensive experiments are conducted to evaluate the proposed method by using both manual-labelled and auto-labelled datasets. The experimental results explicitly demonstrate that our method outperforms state-of-the-art models.

The novelties of our research work are presented as follows: To the best of our knowledge, the proposed method is the first to apply the deep learning-based method to detect public concerns, which rapidly assists the authority to understand people's anxiety and fears about COVID-19; Furthermore, the concern relation is extracted along with concerns, helping to identify any potential social crisis; We are the first to define a concern graph which contributes to the detection of concerns and corresponding relationships, which leads to the performance improvement of the proposed method. Our contributions in this research work are summarised below:

- A concern graph data structure is defined to capture the inherent structural information of concerns more efficiently.
- A novel end-to-end model is presented to jointly extract concerns and relations consisting of Concern Graph (CG) and shared state of concerns.
- The proposed model is evaluated on manual-labelled data and auto-labelled data, and the results indicate the proposed method is effective for auto-labelled data.

2. Related Work

In this section, the existing studies are firstly reviewed, which are related to public concern mining and detection. Then, modern Named Entity Recognition (NER) and Relation Extraction (RE) approaches are inspected and compared since the concern detection, defined in this paper, tends to explore the concern entities and the corresponding relations. Finally, the GCN and its variants are reviewed since GCN has been widely adopted in NER and RE based on recent studies.

2.1. Concern Detection

Social media has become a prevalent platform for people to communicate and express their opinions. With the outbreaks of the pandemic, i.e., Ebola, Zika, COVID-19, how to effectively extract people's opinions and address the public concern in pandemic situations has attracted great attention to researchers. Thus, great efforts have been dedicated to the analysis of public response to pandemics on social media platforms, e.g., Twitter. The current approaches are mainly categorised into two types of methods: probabilistic model-based and deep learning-based. In probabilistic-based models, Latent Dirichlet Allocation (LDA) is commonly used for public concern extractions. For example, Allison et al. apply topic modelling to detect themes of public concern from Ebola tweets and reveal major insights to inform communication strategies (Lazard et al., 2015). Kim et al. conduct content and sentiment analysis on Ebola Twitter (Kim et al., 2016). Five themes are identified from Zika-related Twitter content by Fu et al. through content analysis (Fu et al., 2016). Chandrasekaran et al. conduct a temporal assessment on COVID-19-related tweets to uncover public concern trends by extracting topics and predicting sentiment scores (Chandrasekaran et al., 2020). Xue et al. utilise LDA to analyse public response towards COVID-19 pandemic on the social media platform, aiming to identify popular uni-grams and bi-grams topics from tweets (Xue et al., 2020). Wahbeh et al. adopt a qualitative analysis tool to detect recommendations, topics, and opinions related to the COVID-19 pandemic from Twitter (Wahbeh et al., 2020). Whereas, probabilistic model-based methods perform poorly on public concern identification since contextual information is ignored. By contrast, deep learning-based methods are able to retain contextual features of sentences. Nowadays, deep learning is widely adopted as a popular approach for many Natural Language Processing (NLP) tasks, such as sentiment analysis. By employing such an approach, many studies aim to extract insightful information for assisting the authorities in making appropriate responses and reactions (Wang et al., 2020; Yin et al., 2020; Chen et al., 2020).

However, most existing research works only identify a few pre-defined public concerns, but neglect the relations between the concerns. Without concern relations, it is difficult to identify the cause of public concerns or reveal people's thoughts behind the expressed concerns. Different from the above two types of approaches, our proposed method is able to capture regional and sequential features of a sentence and assist the extraction of public concerns with the corresponding relations.

2.2. Named Entity Recognition

Named Entity Recognition (NER), also referred to as Entity Extraction (ER), is one of the classic tasks of NLP, which aims to identify and classify named entities from unstructured text into pre-defined categories (Mohit, 2014). Recent studies have shown two typical NER approaches, i.e., traditional statistical models and deep learning-based methods. Zhou et al. propose an entity extraction model with a chunk tagger method based on the Hidden Markov Model (HMM), and the model outperforms the hand-crafted rules-based models (Zhou & Su, 2002). Lafferty et al. present Conditional Random Fields (CRF) to segment and label sequence data by building a probabilistic model (Lafferty et al., 2001). However, traditional statistical models perform poorly on complex sentences because they fail to discover hidden features from data. Compared with traditional methods, deep learning-based approaches are able to learn latent representations from raw data and achieve promising performance. Santoso et al. apply Bi-directional Long Short Term Memory (Bi-LSTM) to perform a sequence classification (NER and Part-of-Speech) by understanding the context of the input on the Indonesian language dataset (Santoso et al., 2021). Lample et al. propose a novel neural architecture by relying on character and word representations, which combines Bi-LSTM and CRF (Lample et al., 2016). Similarly, Ma et al. propose a novel deep learning-based model by combining Bi-LSTM, Convolutional Neural Network (CNN), and CRF (Ma & Hovy, 2016). Nowadays, modern state-of-the-art models adopt context-dependent embeddings, e.g., ELMo (Peters et al., 2018), Flair (Akbik et al., 2018), and BERT (Devlin et al., 2019), to encode the input.

Although deep learning-based models are capable of capturing contextual features of data, interaction information between entities is neglected. Different from the above models, apart from contextual information, we also propose a designated Concern Graph (CG) to capture specific features of entities, enabling our method to perform better on Twitter data.

2.3. Relation Extraction

As a fundamental task in the NLP field, Relation Extraction (RE) aims to detect and classify the semantic relationship between entity mentions. Early research works mainly focus on rule-based models, in which proper rules are difficult to define without domain knowledge. To address such an issue, many efforts have been dedicated to kernel-based models with manual-labelled data (Culotta & Sorensen, 2004; Zhou & Zhu, 2011; Seewald & Kleedorfer, 2007). The key weakness of kernel-based methods is that contextual features are not captured, leading to wrong relation extraction on data with a long sentence. Recently, deep neural networks have been applied to relation extraction due to their supremacy in terms of accuracy. Therefore, some popular deep learning models, e.g., CNN, LSTM, and GCN, are utilised to learn contextual features of data and achieve better performance than kernel-based models (Zeng et al., 2014; Miwa & Bansal, 2016; Fu et al., 2019).

Apart from extracting entity and relation separately, many other studies investigate joint methods to extract both simultaneously. For example, Arzoo et al. propose an attention-based RNN model for joint entity mentions and relations extraction (Katiyar & Cardie, 2017). Zheng et al. present a novel tagging strategy to covert sequence labelling and classification tasks to a tagging problem and extract entities and relations directly using the joint model (Zheng et al., 2017). Miwa et al. use Tree-LSTM with bidirectional sequential LSTM to extract entity and relation simultaneously (Miwa & Bansal, 2016). Without any manually extracted features, Bekoulis et al. model NER using a CRF layer and the RE as a multi-head selection problem to extract entities and relations simultaneously (Bekoulis et al., 2018). Zeng et al. propose a sequence-to-sequence model with a copy mechanism to extract entity and relation (Zeng et al., 2018). Hang et al. use one BERT-based parameter-sharing layer to capture the features of entities and relations, then extract entities and relations by applying a source-target BERT model and a three-step overlapping model, respectively (Hang et al., 2021).

However, the existing NER, RE, and joint entity and relation extraction models suffer from two issues. First, existing models only discover contextual features of a sentence and neglect entity features, which is vital for entity extraction. Second, relation extraction mainly relies on a sentence's contextual features, and the information of the corresponding entity relationships is ignored. This can be a severe problem for social media data, where numerous grammatical mistakes exist in sentences. To address these two issues, we combine contextual and concern features for concern identification and integrate learned concern features with the module of relation extraction.

2.4. Graph Convolutional Network

Graph Convolutional Network (GCN) has demonstrated advent advantages in capturing the dependency structure of sentences, and it has been widely adopted in many NLP tasks (Battaglia et al., 2016; Defferrard et al., 2016; Hamilton et al., 2017). As an extension of GCN, Bi-directional Graph Convolutional Network (Bi-GCN) can improve the performance of graph structure data. Figure 1 shows the overview of Bi-GCN. In each hidden layer, the model learns the feature description for the current node and its neighbours, along with the graph structure including both directions from the current node to neighbours and the reversed direction. The output layer can obtain information from backward and forward states simultaneously.

Hong et al. present a joint model based on GCN to perform entity and relation extraction by considering context and syntactic information of sentences (Hong et al., 2020). Zhang et al. utilise GCN over a pruned dependency tree to tackle the relation extraction (Zhang et al., 2018). Inspired by the existing studies, we incorporate GCN into the proposed model to effectively preserve the dependency information of sentences. Furthermore, concern states are integrated with GCN to improve the accuracy of relation extraction.

In this paper, we proposed an end-to-end model with a concern graph module to perform joint extraction of concerns and relations. Meanwhile, we integrate the concern states from Bi-LSTM with the input features of Bi-GCN to enhance the influences from concerns to improve relation extraction performance.

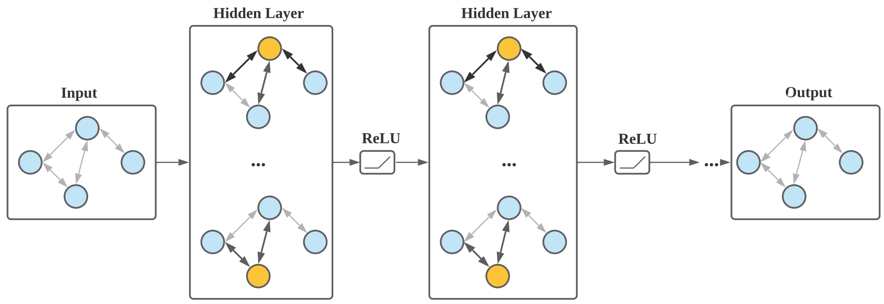

Figure 1: Bidirectional Graph Convolutional Network (Bi-GCN) Overview

3. Preliminaries

In this section, the relevant definitions are presented, including public concerns, concern relations, and graphs. In addition, the concern detection problem is formally formulated.

3.1. Formal Definition

Definition 1: Concern refers to people's worry about a real or imagined issue. Public concern represents a word or a phrase in a tweet towards which most people express strong opinions about a particular aspect of the pandemic. Given a concern set $C = \{c_1, ..., c_n\}$, the ith potential concern detected in tweet t_j can be defined as $c_i^j = (ce_i^j, cs_i^j)$, where $ce_i^j \in CE$ is the concern entity identified in the tweet t_j and it can be words or phrases, e.g., *"China"*, *"corona emergency relief"* and *"florida medical examiner"*. $C = \{c_i^t | i \in [1, N], t \in T\}$ denotes the set of public concerns detected in the Twitter dataset T. For each concern c_i, there is one attribute named type ct_i, where $ct_i \in CT$ and $CT = \{ct_1, ..., ct_n\}$ is the set of concern types. The concern score cs_i^j of concern c_i^j is calculated by Equation 1:

$$cs^j = (1 - \theta) * |sp^j| + \theta * \tilde{rt}^j, \tag{1}$$

where the range of cs_i^j is [0,1], where the greater the value is, the more likely it becomes a concern. $\theta \in [0, 1]$ refers to the weight parameter. $sp_i^j \in [-1, 1]$ denotes the sentiment polarity of the tweet t_j, where -1 indicates an extremely negative attitude, 0 means a neutral attitude, and +1 implies an extremely positive attitude. rt_i^j represents the retweet count of tweet t_j and $\tilde{rt}^j \in [0, 1]$ describes the normalised value of rt^j.

Definition 2: Concern Relation describes the relationship between public concern pairs. We use $r_{m,n}^j \in R$ to present the relation between concern c_m^j and c_n^j in tweet t_j, where $r_{m,n}^j$ is unidirectional relation, i.e., the same as $r_{n,m}^j$, and R is the set of relations extracted from Twitter dataset T.

Definition 3: Concern Triple is the fundamental element of the public concern graph which is extracted from a tweet. To present the real meaning of concern, some short words or phrases in the tweet are very limited in context information. Whereas, the concern triple is capable of semantically representing what concern is about. A public concern triple in the tweet t_j, $ct^j_{m,n} = (s^j_m, r^j_{m,n}, o^j_n)$, has three components, i.e., s^j_m, $r^j_{m,n}$ and o^j_n, referring to as the subject, relation, and object of the concern triple, respectively. The s^j_m and o^j_n are extracted entities, and $r^j_{m,n}$ is the extracted relation based on dependency parser analysis of the tweet t_j.

Definition 4: Concern Graph aims to explore discriminative public concerns and what kind of relations exist between concerns. To present the relation of public concerns, the Concern Graph (CG) is proposed as the control signal for capturing public concerns. CG of the tweet t_j can be denoted as $G = (\nu, \varepsilon)$, where ν is the set of nodes, and ε is the edges set. As shown in Figure 2, nodes of CG are classified into four categories: (1) object o^j, subject node s^j; (2) relation note r^j; (3) attribute node a^j including concern type ct^j; (4) concern score cs^j.

The CG G is constructed via the following steps:

1. Detect public concern c^j_i and add it to G, where c^j_i is grounded in the tweet t_j.
2. Extract the descriptive details of concern c^j_i as the attribute $a^j_{i,l}$ including type ct^j_i and score cs^j_i, then add them to G and assign an un-directed edge from c^j_i to a^j_{il}, where $|l|$ is the number of attributes towards concern c^j_i.
3. Identify the relation r_{ik} between concerns c^j_i (subject in concern triple) and c^j_k (object in concern triple), which is a unidirectional type of relation, adding relation node r_{ik} to G and assigning edges from c^j_i to r_{ik} and from r_{ik} to c^j_k.

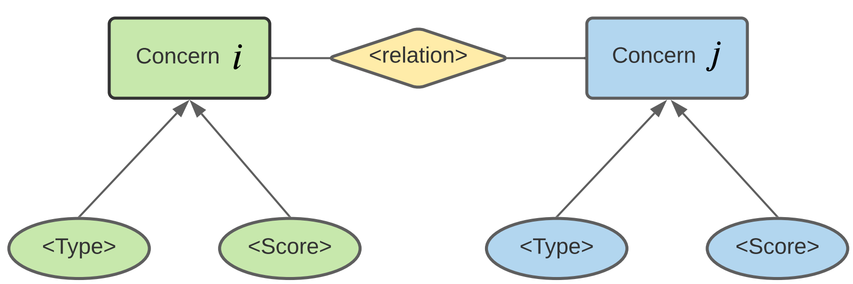

Figure 2: Concern Graph: each concern has two attributes, i.e. type and score, along with relation to another concern to form the concern graph.

3.2. Problem Formulation

In the previous section, the related definitions are described. Based on the definitions, our model aims to jointly extract typical concerns $\{c_i^j | c_i^j \in C\}$ from tweet t_j and concern relations $\{r_{mn}^j | r_{mn}^j \in R\}$, where r_{mn} is the relation between concern c_m and c_n from Twitter dataset T by constructing CG.

4. Graph-based Concern and Relation Extraction

For a set of tweets T, the goal of our method is to identify public concerns $C = \{c_1, ..., c_n\}$ and concern relations $R = \{r_1, ..., r_n\}$. In this section, the joint extraction of concerns and relations model with concern graph is illustrated in Figure 3. The proposed method consists of four main components, i.e., embedding layer, encoding layer, concern decoding layer, and concern relation extraction layer. Each component is described in detail below. The embedding layer is introduced in Section 4.1, followed by the encoding layer in Section 4.2. Concern decoding and concern relation extraction layer are presented in Sections 4.3 and 4.4, respectively. The model objective function is explained in Section 4.5.

The proposed method is named Concern Graph-based Concern and Relation Extraction (CG-CRE). In Algorithm 1, the training process of CG-CRE is demonstrated for improved understanding. All weight parameters are initialised in Bi-LSTM and Bi-GCN. Subsequently, the CG embedding is generated in each epoch and then computes the loss function of concern and relation. The final objective function for model training is calculated based on concern and relation loss function with a trade-off coefficient.

4.1. Embedding Layer

Since deep learning models are integrated into the proposed method, word tokens and proposed CG need to be transformed into low-dimensional vectors by the embedding layer. Given a tweet $t = \{w_1, ..., w_i, ..., w_n\}$, where w_i denotes the ith word in the tweet, pre-trained BERT model is used to generate word embedding set $\grave{X} = \{\tilde{e_1}, ... \tilde{e_i}, ..., \tilde{e_n} \,|\, \tilde{e_i} \in \mathbb{R}^d\}$, where $\tilde{e_i}$ represents the embedding of word w_i and d means the embedding dimension.

To enhance model input features, we further encode proposed CG G to obtain CG node embedding $\acute{x}_i^{(0)}$ in Equation 2:

$$\acute{x}_i^{(0)} = \begin{cases} (v_i^{(dep)} + v_i^{(pos)}) \odot W_{cr}[0], & if\, i \in C; \\ v_i^a \odot W_{cr}[1], & if\, i \in A; \\ v_i^r \odot W_{cr}[2], & if\, i \in R; \end{cases} \tag{2}$$

where $v_i^{(dep)}$ and $v_i^{(pos)}$ denote the syntactic dependency relation and POS tag feature, respectively. Both $v_i^{(dep)}$ and $v_i^{(pos)}$ are used to capture the meaning of tweet and words syntactic dependency. C represents the concern set. v_i^a represents the attribute features, including concern type and score. A means

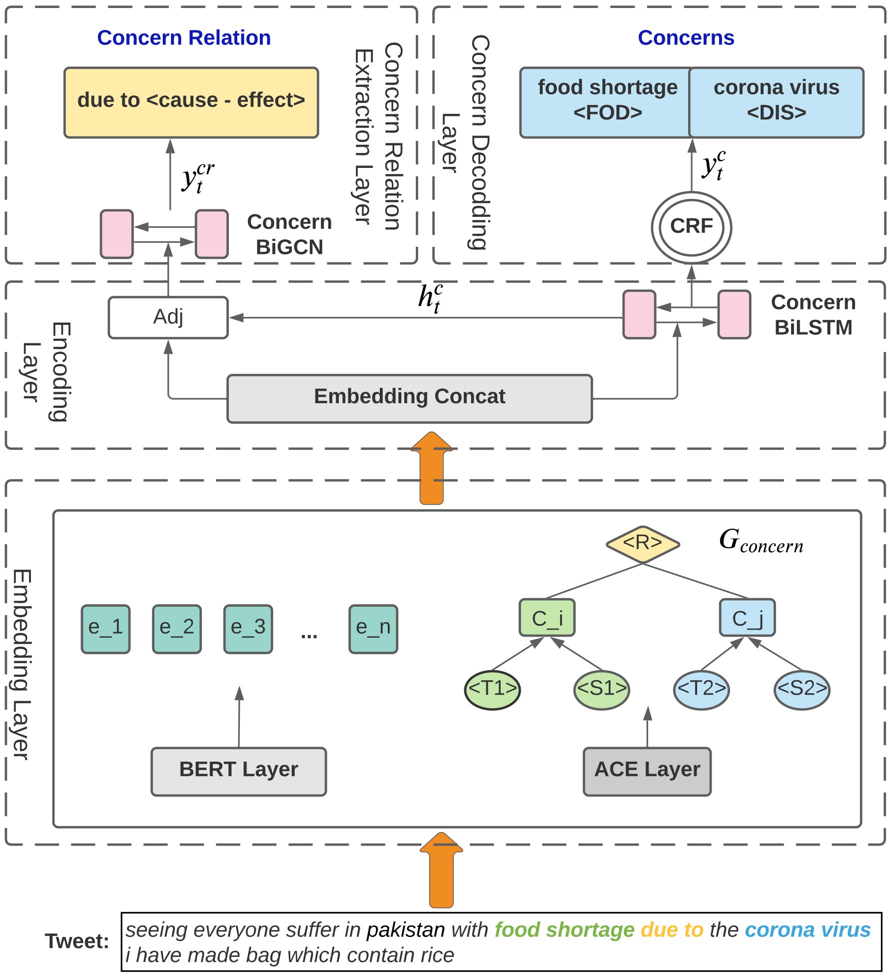

Figure 3: The overview of CG-CRE model.

Algorithm 1: Training Process of CG-CRE Model

Input : T, υ_w, υ_{cg}, E, B, lr, d
T indicates labelled Twitter corpus for model training
υ_w represents BERT embedding of Twitter corpus
υ_{cg} is embedding of CG
EP is epoch number
B is batch size
lr indicates learning rate
d is embedding dimension

Output: L
L is the loss function value

initialization: weight vector W;
while *ep in EP* **do**
  while *b in B* **do**
    generate embedding $\acute{x}_i$;
    compute concern hidden state ${h_t}^{(b)}$ as Equation 8 ;
    $L^{(b)}_{(c)} = max(\sum log(P^{(b)}_{(c)} = S^{(b)}_{(c)}))$;
    $L^{(b)}_{(r)} = max(\sum log(P^{(b)}_{(r)} = S^{(b)}_{(r)}))$;
    $L^{(b)} = L^{(b)}_{(c)} + \alpha * L^{(b)}_{(r)}$;
  end
end

attribute set. v_i^r indicates relation feature, and R is relation set. $W(\cdot) \in \mathbb{R}^{3\times d}$ refers to parameters, where d means the feature dimension.

As illustrated in Figure 4, the input tweet is split into words that are used to generated inputs for BERT layer and embedding layer. The input words are encoded into positional embeddings, segment embeddings, and token embeddings, which are essential parts of the attention mechanism of BERT. After dependency relation vector, POS tag vector, attribute vector, and relation vector are encoded, they are fed into the embedding layer to generate CG embeddings. The final output representation is computed by concatenating the outputs from BERT layer and embedding layer.

4.2. Encoding Layer

To capture long-distance dependencies and forward and backward features between tokens in tweets, Bi-LSTM is used in this paper. The Bi-LSTM contains forward and backward layers, and a concatenation layer of backward and forward state information. The embeddings in Section 4.1 are concatenated as the input of the concern encoder layer. The Bi-LSTM encoding layer is defined by using Equations 3 - 8:

$$i_t = \sigma(W^{(i)}_{ex} * [\tilde{e}_i; \acute{x}^{(i)}_j] + W^{(i)}_h * h_{t-1} + b^{(i)}) \tag{3}$$

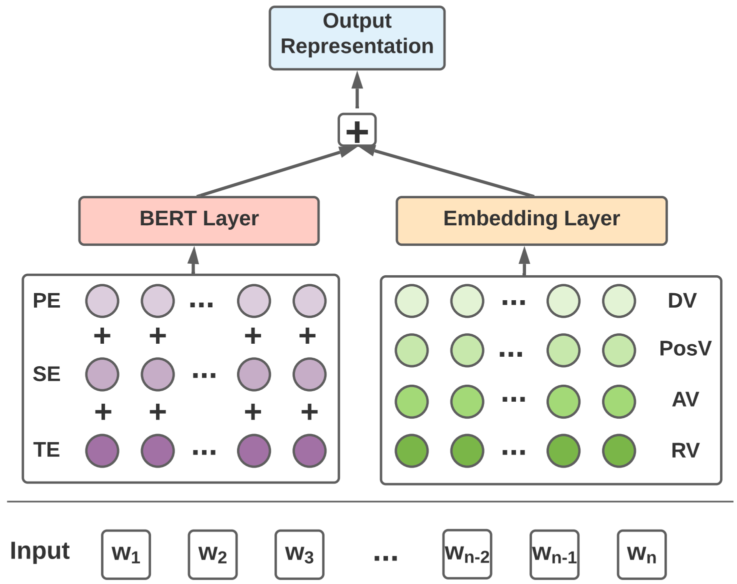

Figure 4: The embedding of CG-CRE model. PE, SE, and TE mean positional embeddings, segment embeddings, and token embeddings, respectively. DV, PosV, AV, and RV refer to dependency relation vector, POS tag vector, attribute vector, and relation vector, respectively.

$$f_t = \sigma(W_{ex}^{(f)} * [\tilde{e}_i; \acute{x}_j^{(f)}] + W_h^{(f)} * h_{t-1} + b^{(f)}) \tag{4}$$

$$o_t = \sigma(W_{ex}^{(o)} * [\tilde{e}_i; \acute{x}_j^{(o)}] + W_h^{(o)} * h_{t-1} + b^{(o)}) \tag{5}$$

$$u_t = \sigma(W_{ex}^{(u)} * [\tilde{e}_i; \acute{x}_j^{(u)}] + W_h^{(u)} * h_{t-1} + b^{(u)}) \tag{6}$$

$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \tag{7}$$

$$h_t = o_t \odot tanh(c_t), \tag{8}$$

where σ is sigmoid activation function, $W(\cdot)$ refers to weight parameters, and $[;]$ is a vector concatenation operation. $\tilde{e}_i$ and $\acute{x}_j(\cdot)$ denote word embedding and embedding of CG G defined in Section 4.1. In Equations 3 - 6, $b(\cdot)$ refers to the bias vector, and $\odot$ represents element-wise multiplication. c and h denote cell state and hidden state, respectively, carrying information from the previous layer to the next layer. Because Bi-LSTM is applied in our method, the hidden state is obtained by concatenating hidden states in both directions, namely, forward direction $\overrightarrow{h_t}'$ and backward direction $\overleftarrow{h_t}'$, therefore, the final hidden state can be denoted as $h_t' = [\overrightarrow{h_t'}, \overleftarrow{h_t'}]$. By passing hidden state to a fully connected neural network, the final output of Bi-LSTM can be defined in Equation 9:

$$O = W^o * h_t' + b^o, \tag{9}$$

where W^o is the output weight parameters and b^o is the bias vector.

4.3. Concern Decoding Layer

In the proposed model, the CRF is employed to produce a tag sequence since it can produce a higher tagging accuracy than that of the existing models (Hong et al., 2020). For one tweet $t = \{w_1, ..., w_n\}$, the goal is to predict the concern tag sequence $Y^{(c)} = \{y_1^{(c)}, y_2^{(c)}, ..., y_n^{(c)}\}$ where n denotes the number of words and superscript (c) means the notation of concern. Thus, the CRF score can be defined as in Equation 10:

$$S^{(c)}(t, Y^{(c)}) = \sum_{i=1}^{n} O_{i,y_i^{(c)}} + \sum_{i=1}^{n} T_{y_i^{(c)},y_{i+1}^{(c)}}, \tag{10}$$

where $O \in \mathbb{R}^{n\times k}$ indicates the matrix of scores output from the previous encoding layer with k as the number of distinct tags, and $O_{i,j}$ denotes the score of the jth tag of the ith word in tweet t. T represents a matrix of transition scores as being introduced in (Huang et al., 2015), and $T_{i,j}$ means the score of a transition from tag i to tag j. Then, for input tweet t, the probability of a given sequence of tags over the sequence of predicted tags $Y^{(c)}$ is defined by applying the Softmax layer as in Equation 11:

$$P^{(c)} = \frac{e^{S^{(c)}(t,Y^{(c)})}}{\sum_{\widetilde{y}^{(c)} \in Y_X^{(c)}} e^{S^{(c)}(t,\widetilde{y}^{(c)})}}, \tag{11}$$

In Equation 11, $Y_X^{(c)}$ denotes all possible concern tag sequences for tweet t.

4.4. Concern Relation Extraction Layer

Given concern set $C = \{c_1, ..., c_m\}$ in tweet $t = \{t_1, ..., t_n\}$, the goal is to extract corresponding relation $r_i \in R$. Except for sequential features, Bi-GCN is utilised to capture regional features from the tweets. Both forward and backward directions are considered and the hidden state of Bi-GCN is defined using Equations 12 - 14:

$$\overrightarrow{h_t''} = \varsigma(\sum_{v \in \overrightarrow{N(w)}} (\overrightarrow{W_h} * [\overrightarrow{h_{t-1}^{v}{''}}; \overrightarrow{h_{t-1}{'}}] + \overrightarrow{b})) \tag{12}$$

$$\overleftarrow{h_t''} = \varsigma(\sum_{v \in \overleftarrow{N(w)}} (\overleftarrow{W_h} * [\overleftarrow{h_{t-1}^{v}{''}}; \overleftarrow{h_{t-1}{'}}] + \overleftarrow{b})) \tag{13}$$

$$h_t'' = [\overrightarrow{h_t''}; \overleftarrow{h_t''}], \tag{14}$$

where ς represents ReLU activation function, h_t'' refers to the hidden state at tth layer and $\overrightarrow{h_{t-1}}'$ indicates the shared hidden state from the concern detection module. $\overrightarrow{N(w)}$ describes the neighbours of word w in the forward direction and $\overleftarrow{N(w)}$ means the neighbours of word w in the backward direction. $\overrightarrow{W_h}$ and $\overleftarrow{W_h}$ represent weight parameters in the forward and backward direction, respectively. $\overrightarrow{b}$ and $\overleftarrow{b}$ are the bias of the model. h^t refers to the final hidden state of word w, concatenating hidden states in both directions.

By using hidden states of Bi-GCN, the relation tendency score $S^{(r)}_{(r_{ij}|c_i,c_j)}$ is defined in Equation 15:

$$S^{(r)}_{(r_{ij}|c_i,c_j)} = W^{(r)} * \varsigma(W^{(r)}_{c_i} * h''_{c_i} + W^{(r)}_{c_j} * h''_{c_j} + b^{(r)}), \tag{15}$$

where superscript (r) means the notation of concern relation. $S^{(r)}_{(r_{ij}|c_i,c_j)}$ represents the tendency score of concern relation on concerns pair (c_i, c_j). $W^{(r)}$, $W^{(r)}_{c_i}$ and $W^{(r)}_{c_j}$ are weight parameters. $b^{(r)}$ denotes the bias term. The activation function (Softmax) is applied to the tendency score $S^{(r)}_{(r_{ij}|c_i,c_j)}$ to obtain the probability of relation $r_{i,j}$ in Equation 16:

$$P^{(r)} = \sigma(S^{(r)}_{(r_{ij}|c_i,c_j)}) \tag{16}$$

4.5. Model Objective Function

In this subsection, the final objective function for model training is described. To train the proposed model, the maximum log-likelihood is used as the loss function and maximise combined loss functions of concern and relation by using Equations 17, 18, and 19:

$$L_{(c)} = max(\sum_{i=1}^{|\mathbb{R}_T|}\sum_{w=1}^{|W_i|} log(P^{(c)}_w = S^{(c)}_w|t_i, \Theta)) \tag{17}$$

$$L_{(r)} = max(\sum_{j=1}^{|\mathbb{R}_T|} log(P^{(r)}_w = S^{(r)}_w|t_j, \Theta)) \tag{18}$$

$$L = L_{(c)} + \alpha * L_{(r)} \tag{19}$$

where $|\mathbb{R}_T|$ is the size of the training dataset, t_i and t_j is the ith and jth tweet in the training dataset, respectively. $|W_i|$ is the sentence length. $\alpha \in [0, 1]$ is a trade-off coefficient between loss of concern and concern relation, and the larger value means the greater influence of concern relation on the proposed method.

Table 1: Statistics of the Concern Categories

Type	Tweets	Concern Category							
		FIN	GOV	DIS	MED	PER	LOC	FOD	DAT
Manual-labelled	1761	315	457	1239	471	289	341	204	206
		9%	13%	35%	13%	8%	10%	6%	6%
Auto-labelled	40068	4341	19941	23853	6944	10977	1519	1498	11063
		5%	25%	30%	9%	13%	2%	2%	14%

5. Experiments

In this section, extensive experiments are conducted to evaluate the proposed approach by using COVID-19 Twitter datasets. First, COVID-19 dataset collection and pre-processing are described. Second, the proposed approach is compared against six state-of-the-art baselines in terms of precision, recall, and F1 score. Third, quantitative analytical results and conduct ablation studies are presented following the experimental results. Finally, a case study is given to illustrate the effectiveness of our approach.

5.1. Dataset and Experiment Setting

Twitter is one of the largest social media platforms, providing a rich source for evidence. It is easy for people to obtain the tweets associated with COVID-19 by using API. The experiments are conducted by using a public large-scale Twitter dataset about COVID-19, which contains English language-specific tweets from 204 different countries and territories (Lamsal, 2021). The dataset is proposed in the scientific literature for research with topics related to COVID-19.

The dataset has been pre-processed in two ways, i.e., manual annotated and auto-annotated. In the former, the annotators label the tweets according to the concern definitions and formulations. While, in the latter, tweets are annotated by using the approach proposed in our past research work (Shi et al., 2021).

Many prior research works have explored people's reactions and attempt to discovered wide-spreading topics about COVID-19 (Li et al., 2020; Killeen et al., 2020; Hou et al., 2020; Kaveh-Yazdy & Zarifzadeh, 2020; Li et al., 2020). Based on the findings and conclusion of these works, the most popular topics are extracted and eight types of concerns are defined, i.e., Finance (FIN), Government (GOV), Disease (DIS), Medicine (MED), Person (PER), Location (LOC), Food (FOD), and Date and Time (DAT). On top of that, two types of relations among the concerns, i.e., co-occurrence and cause-effect, are investigated. This is because both types of relations are capable of capturing implicit information about public concerns, demonstrating their associations and potential causes. For instance, by analysing the tweet "... due to the locked transportation..., farmers forced to dump green chilli ...", it is important to know the concern "green chilli" is dumped due to the concern "locked transportation" in the time of COVID-19 pandemic. The statistics of concerns and the relations are listed in Table 1 and Table 2, respectively.

Table 2: Statistics of Concern Relation Categories

Type	Tweets	Concern Relation	
		CO_OCC	CA_EFF
Manual-labelled	1761	932 53%	829 47%
Auto-labelled	40068	19485 49%	20583 51%

Table 3: Statistics of the manual-labelled and auto-labelled dataset

	Train	Test
Manual-labelled	1418	343
Auto-labelled	32264	7804

The statistics of datasets are listed in Table 3. The dataset is divided into 2 sub-datasets: train dataset and test dataset, occupying 80% and 20%, respectively.

5.1.1. Evaluation Metrics

In this paper, three standard evaluation metrics, i.e., precision, recall, and F1 score, are employed to evaluate the proposed model.

The outcome of predicted concerns is considered correct only when both of the concerns in one tweet are predicted correctly. In other words, $(c1, c2)$ is recognised as a correct concern pair if $c1$ and $c2$ are correctly predicted at the same time. Correspondingly, the relation prediction is considered valid only when the associated concern pair is correctly predicted.

5.1.2. Hyper-parameters

The language model BERT has been proven to be effective for many natural language processing tasks. In the experiments, the pre-trained BERT-base [1] is utilised to obtain word representations of tweet corpus, and the hidden dimension of embedding H is set as 768. Because BERT uses WordPiece tokenizer to generate word tokens, some concern words may break into several pieces. To detect concern in one word instead of sub-word pieces, the corresponding representations of sub-word tokens are averaged to get one concern representations. For example, the representation of the concern "covid-19" is the average of representations of three-word pieces "co", "##vid", "##19". Our network is regularised by using dropout at the embedding layer, with a dropout ratio of 0.2. Bi-LSTM and GCN are adopted as the encoding layer, with 300 LSTM units. We employ the full dependency tree of sentences as the adjacency matrix of GCN.

[1] https://github.com/huggingface/pytorch-pretrained-BERT

5.2. Baselines

The proposed approach is evaluated by comparing against the following baselines.

- ***Joint Model*** (Zheng et al., 2017) is a joint extraction method to detect both entity and relation in one tweet by using a novel tagging scheme. It is an end-to-end model consisting of a Bi-LSTM encoder layer and an LSTM decoder layer.
- ***Copy Mechanism Model*** (Zeng et al., 2018) is a state-of-the-art model for jointly extracting relation triplets from a sentence. It is also an end-to-end model based on seq-to-seq learning with a decoder layer, having two different decoding methods, i.e., one-decoder and multi-decoder. Both different strategies are used as counterparts in the experiments.
- ***SPTree***(Miwa & Bansal, 2016) is a novel end-to-end recurrent neural network model aiming at extracting entities and relations by capturing word sequence and dependency tree substructure features. The stacked bidirectional tree-structured LSTM-RNN models are applied on sequential Bi-LSTM-RNN models to detect both entities and relations with shared parameters.
- ***JointER***(Yu et al., 2020) is a joint entity and relation extraction model which can address the limitations, including redundant entity pairs and ignoring the important inner structure of entities. The model decomposes a joint extraction task into Head-Entity (HE) extraction and Tail-Entity-Relation (TER) extraction to detect head-entity, tail-entity, and relations.
- ***SPERT***(Eberts & Ulges, 2020) is introduced as a span-based model, which can jointly extract entity and relation by conducting lightweight reasoning on BERT embedding and relation classification based on localised and marker-free context features.
- ***CopyMTL***(Zeng et al., 2020) is a multi-task learning framework with copy mechanisms to predict multi-token entities and relations. It is an extremely effective model which can address two existing problems of entity and relation extraction: (1) inaccurate entity extraction caused by failing to differ the head and tail entity; (2) failing to predict multi-token entities.

5.3. Experimental Results and Model Analysis

In this section, we present and analyse the strengths and weaknesses of the proposed method by comparing it against the state-of-the-art models mentioned previously. To ensure the fairness and rationality of the experiments, we select all the counterparts, which incorporate a Bi-LSTM encoder layer.

The experimental results are demonstrated in Table 4, which presents the predicted outcomes, i.e., Precision, Recall, and F1, of the proposed approach as well as the state-of-the-art methods on manual-labelled and auto-labelled

Table 4: Evaluation results of different models on COVID-19 Tweets Dataset

Model	Manual-labelled Tweets			Auto-labelled Tweets		
	Precision	Recall	F1	Precision	Recall	F1
One-Decoder	0.167	0.161	0.164	0.328	0.321	0.326
Multi-Decoder	0.159	0.152	0.156	0.399	0.347	0.373
NovelTagging	0.273	0.336	0.302	0.570	0.593	0.582
SPTree	0.424	0.349	0.383	0.434	0.366	0.397
JointER	**0.644**	0.369	0.469	0.405	0.314	0.354
SPERT	0.239	**0.675**	0.339	0.310	**0.839**	0.421
CopyMTL-One	0.427	0.393	0.412	0.461	0.413	0.447
CopyMTL-Mul	0.538	0.515	0.530	0.594	0.551	0.573
Proposed Model	0.545	0.630	**0.567**	**0.638**	0.642	**0.592**

datasets. As can be observed from the table, the proposed approach outperforms the others in terms of F1 score, which proves its effectiveness. Specifically, in Figures 5 - 7, the CG-based model outperforms One-decoder, Multi-decoder, NovelTagging, SPTree, and CopyMTL models on both manual-labelled and auto-labelled Twitter datasets. Although JointER and SPERT achieve better performance than that of ours in terms of precision and recall on the manual-labelled dataset, SPERT leverages the pre-trained BERT model to obtain contextual features of sentences, but the inner structure of entities is neglected, which inevitably hinders the performance of entity and relation extraction. The embeddings in JointER are initialised using the shallow representatives model, i.e., Glove (Pennington et al., 2014), without context-specific information, which is critical for entity and relation extraction models.

The promising performance of the proposed approach mainly attributes to its structural design. First, the interaction of the CG structure captures the inner dependency between concerns. Second, the shared state passing from concern extraction module to relation extraction module, provides important concern features for relation extraction. It is worth noting that baselines can achieve state-of-the-art results on high-quality datasets, e.g., NYT and WebNLG, but the performance significantly degrades on the noisy and imbalanced social media data. The grammatical mistakes of tweets make it difficult to capture relations between concerns. NovelTagging and SPTree utilise novel tagging but cannot carry out promising results. Other baselines, including One-Decoder, Multi-Decoder, JointER, and CopyMTL, apply Bi-LSTM to capture sequential features of concerns, but they fail to detect the relation features and concerns due to the unstructured sentences in the tweet dataset.

To better understand the experimental results, some examples are presented, which are obtained by applying the proposed method to COVID-19 tweets. The examples are demonstrated in Figure 8. The proposed method can detect two concerns (e.g., "food shortage", "corona virus") and the concern types (e.g., "FOD", "DIS"). Moreover, the relation (e.g., "CA_EFF") between concerns is further extracted from the tweets. Incorporating with concern relation, the

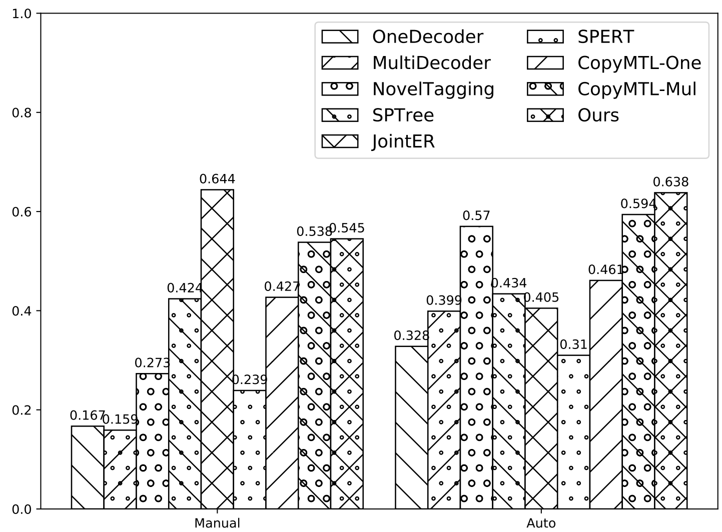

Figure 5: Experiment results (Precision) on COVID-19 tweets dataset.

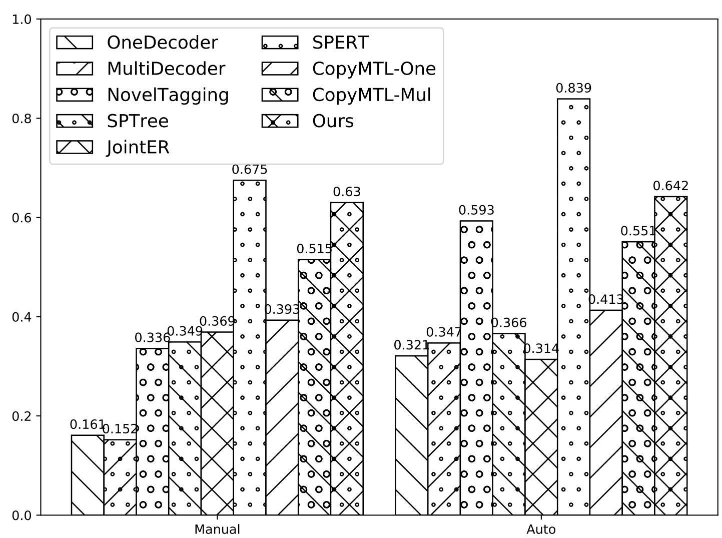

Figure 6: Experiment results (Recall) on COVID-19 tweets dataset.

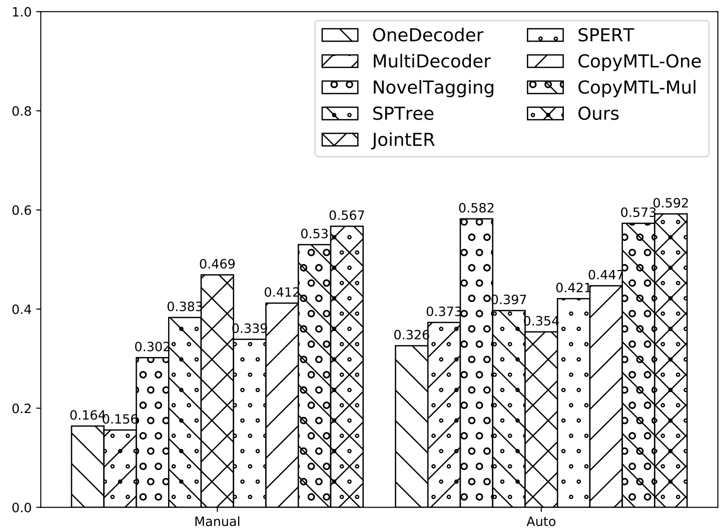

Figure 7: Experiment results (F1 score) on COVID-19 tweets dataset.

proposed method boost in reveal meaningful information of public concerns instead of only knowing isolated concerns from tweets.

To study the effect of training and test data distribution on the proposed model performance, some experiments are conducted by dividing tweets into training and testing data with different training-testing ratios. Both manual and auto labelled data are divided into three groups with three types of training-testing ratios, 70%-30%, 80%-20%, and 90%-10%. The experimental results are summarized in Table 5. As can be seen, the proposed method achieves the best performance on manual and auto labelled tweets when datasets are split into 80% training and 20% test data. when the training data is increased to

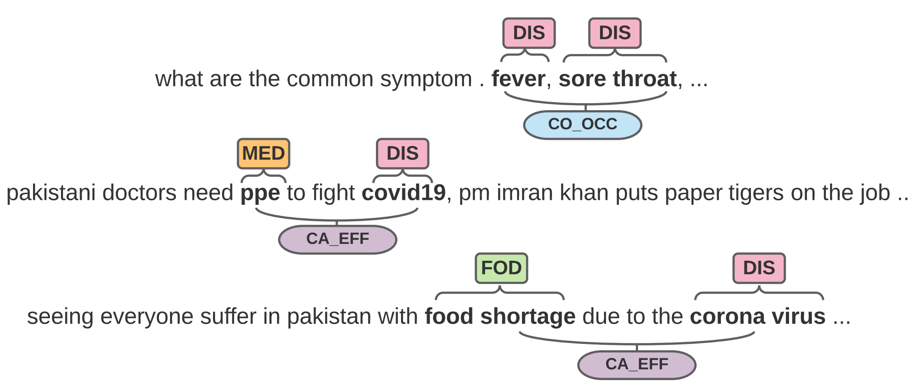

Figure 8: Examples of the proposed method on the COVID-19 tweets.

Table 5: The performance of the proposed method with different training and testing data distribution

Test-Train (%)	Manual-labelled Tweets			Auto-labelled Tweets		
	Precision	Recall	F1	Precision	Recall	F1
30-70	0.497	0.573	0.488	0.569	0.586	0.536
20-80	0.545	0.630	0.567	0.638	0.642	0.592
10-90	0.533	0.614	0.550	0.621	0.618	0.571

Table 6: The performance of the proposed method with different embedding dropout values

Embedding Dropout	Manual-labelled Tweets			Auto-labelled Tweets		
	Precision	Recall	F1	Precision	Recall	F1
0	0.540	0.612	0.546	0.618	0.642	0.573
0.1	0.538	0.615	0.554	0.623	0.644	0.579
0.2	0.533	0.626	0.561	0.629	0.650	0.587
0.3	0.545	0.630	0.567	0.638	0.642	0.592
0.4	0.524	0.602	0.553	0.611	0.625	0.581

90% of total tweets, the proposed method begins overfitting, and the F1 score drops by 2.1%. The increasing of training data can be helpful in increasing the performance of the proposed method, while it can also decrease the proposed method performance due to the overfitting of the model.

In Table 4, the proposed method is compared against other models which can detect entities and extract relations simultaneously, and the experimental results demonstrate the superiority of the proposed method in concern and relation extraction. Moreover, we explore the effect of the embedding dropout on the performance of the proposed method in Table 6. The experimental results show that the performance (e.g., F1 score) will be decreased with low embedding dropout values on both manual and auto labelled datasets, for example, the F1 score decreases 2.1% and 1.9% with embedding dropout 0 and 0.3 on both datasets, respectively. The results also report that the performance improvement is not able to obtain when dropout values are reduced to 0.2 and 0.3.

5.4. Ablation Study

The ablation study in this section aims to investigate the impact of CG and shared state components in the proposed approach.

Since manual labelling a large-scale dataset turns out to be a tedious and non-trivial task, sufficient manual-labelled training data sets are usually not available to conduct public concern extraction and analysis for an emergency event. Furthermore, the public concern coverage in datasets also appears imbalanced, which prevents the existing models from generalisation, subsequently impacting the performance to a large extent. The proposed approach can mitigate this issue, giving an outstanding performance on both manual-labelled and auto-labelled datasets.

Table 7: Ablation study of CG-CRE model on manual-labelled and auto-labelled Tweets Dataset

Dataset	Method	Precision	Recall	F1
Manual-labelled Tweets	CG-CRE (without CG)	0.416	0.482	0.457
	CG-CRE (without shared state)	0.463	0.516	0.494
	CG-CRE (with all components)	**0.545**	**0.630**	**0.567**
Auto-labelled Tweets	CG-CRE (without CG)	0.551	0.583	0.536
	CG-CRE (without shared state)	0.615	0.624	0.586
	CG-CRE (with all components)	**0.638**	**0.642**	**0.592**

Table 7 lists the results of the ablation study. The approach has been reevaluated by comparing the performance against that without CG component and shared state components. It can be seen from the table that, in manual-labelled dataset, CG-CRE with CG and shared state outperforms the models without CG and shared state by 11% and 7%, respectively. While in auto-labelled datasets, it surpasses 6% and 1%, respectively. The results explicitly reveal that CG and shared state components play a significant role in jointly identifying concerns and relations.

5.5. Case Study

In this section, we conduct case studies, presenting some representative public concern extraction examples, to further prove the effectiveness and validity of the proposed approach. Table 8 shows the outputs from three models, including NovelTagging, JointER, and the proposed CG-CRE. In the first case, both concerns and concern relation are identified incorrectly by NovelTagging, and JointER predicts nothing. By contrast, CG-CRE can extract both concerns correctly. Similar outputs are presented in the fifth and the sixth case. As for the second and third cases, NovelTagging only detects one concern correctly and cannot extract the second concern and relation. However, JointER and CG-CRE can accurately identify concerns and concern relations. JointER is not able to carry out the prediction results. In the fourth case, NovelTagging can identify only one concern correctly. JointER is able to obtain accurate predictions, but still remains to be improved in eliminating null prediction. NovelTagging is weak at extracting relations from Twitter datasets.

Based on the experimental results and case studies, we can conclude that the proposed CG-CRE model can yield better performance on both entity recognition and relation extraction than the state-of-the-art models.

6. Conclusion and Future Work

In this paper, an end-to-end model is presented to simultaneously extract concern and concern relations from the social media dataset of COVID-19. GCN and Bi-LSTM are jointly combined to learn sequential and regional dependency features from tweets. In order to capture more features of model input, the influence of graph structure for concern and relation extraction is explored.

Table 8: Outputs from different models on tweets. "pred:[]" means the model predicts null for this tweet. NovelTagging only predicts "c1" and "c2" without concern types.

Models	Tweet
NovelTagging	[**seeing everyone**]$_{c1,r:co\_occ}$ suffer [**in pakistan**]$_{c2,r:co\_occ}$ with food shortage due to the corona virus i have made bag which contain rice
JointER	seeing everyone suffer in pakistan with food shortage due to the corona virus i have made bag which contain rice [**pred:[]**]
CG-CRE	seeing everyone suffer in pakistan with [**food shortage**]$_{c1:FOD,r:ca\_eff}$ due to the [**corona virus**]$_{c2:DIS,r:ca\_eff}$ i have made bag which contain rice
NovelTagging	a greeting from the heart to [**doctors**]$_{c1,r:co\_occ}$, nurses, [**paramedics**]$_{c2,r:co\_occ}$, ... who stand together to tackle the corona epidemic.
JointER	a greeting from the heart to [**doctors**]$_{c1:MED,r:co\_occ}$, [**nurses**]$_{c2:MED,r:co\_occ}$, paramedics, ... who stand together to tackle the corona epidemic.
CG-CRE	a greeting from the heart to [**doctors**]$_{c1:MED,r:co\_occ}$,[**nurses**]$_{c2:MED,r:co\_occ}$, paramedics, ... who stand together to tackle the corona epidemic.
NovelTagging	[**coronavirus**]$_{c1,r:co\_occ}$ could double number of people going hungry. the risk of major interruptions to [**food supplies**]$_{c21,r:co\_occ}$ over the coming months is growing.
JointER	[**coronavirus**]$_{c1:DIS,r:ca\_eff}$ could double number of people [**going hungry**]$_{c2:FOD,r:ca\_eff}$. the risk of major interruptions to food supplies over the coming months is growing.
CG-CRE	[**coronavirus**]$_{c1:DIS,r:ca\_eff}$ could double number of people [**going hungry**]$_{c2:FOD,r:ca\_eff}$. the risk of major interruptions to food supplies over the coming months is growing.
NovelTagging	breaking one of somalia 's greatest artist ha [**died**]$_{c1,r:co\_occ}$ in london after contracting [**corona virus**]$_{c2,r:co\_occ}$ ...
JointER	breaking one of somalia 's greatest artist ha died in london after contracting corona virus ... [**pred:[]**]
CG-CRE	breaking one of somalia 's greatest [**artist**]$_{c1:PER,r:ca\_eff}$ ha died in london after contracting [**corona virus**]$_{c2:DIS,r:ca\_eff}$ ...
NovelTagging	what are the [**common**]$_{c1,r:co\_occ}$ [**symptom**]$_{c2,r:co\_occ}$. fever , sore throat ...
JointER	what are the common symptom . fever , sore throat ... [**pred:[]**]
CG-CRE	what are the common symptom . [**fever**]$_{c1:DIS,r:co\_occ}$, [**sore throat**]$_{c2:DIS,r:co\_occ}$ ...
NovelTagging	social distancing, stay home, [**naija people**]$_{c1,r:co\_occ}$ will not hear. this corona thing has just started with us in this [**country**]$_{c2,r:co\_occ}$, we ...
JointER	social distancing, stay home, naija people will not hear. this corona thing has just started with us in this country, we ... [**pred:[]**]
CG-CRE	[**social distancing**]$_{c1:GOV,r:co\_occ}$, [**stay home**]$_{c2:GOV,r:co\_occ}$, naija people will not hear. this corona thing has just started with us in this country, we ...

The sequential and regional features from the dataset are concatenated, enabling the embedding vectors to represent rich contextual information of both concerns and relations. The proposed model is evaluated on manual-labelled and auto-labelled datasets. The experimental results show that the proposed model can outperform the existing entity and relation extraction models, which demonstrate the effectiveness of the proposed method. Furthermore, the previous methods only work on hand-crafted datasets, while the proposed model turns out to be applicable to both manual-labelled and auto-labelled datasets. Therefore, the proposed method can be easily transferred and applied to other pandemics situations, e.g., Zika, Dengue Fever, and Yellow Fever.

In the future, the plan is to improve the proposed model from two aspects. First, more concern types and concern relation types can be predicted to better understand people's attention and the relation between them. In addition, the time factor can be used to track the trend of one specific concern over time.

7. Acknowledgements

This work was supported by the Callaghan Innovation [CSITR1902, 2020], New Zealand's Innovation Agency.

References

Akbik, A., Blythe, D., & Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 1638–1649).

Battaglia, P., Pascanu, R., Lai, M., Rezende, D. J., & kavukcuoglu, K. (2016). Interaction networks for learning about objects, relations and physics. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (pp. 4509–4517).

Bekoulis, G., Deleu, J., Demeester, T., & Develder, C. (2018). Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, *114*, 34–45.

Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020). Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. *Journal of Medical Internet Research*, *22*, e22624.

Chen, L., Lyu, H., Yang, T., Wang, Y., & Luo, J. (2020). In the eyes of the beholder: Sentiment and topic analyses on social media use of neutral and controversial terms for COVID-19. ArXiv preprint arXiv:2004.10225.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PloS one*, *5*, e14118.

Culotta, A., & Sorensen, J. (2004). Dependency tree kernels for relation extraction. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics* (pp. 423–429).

Damiano, A., & Catellier JR, A. (2020). A content analysis of coronavirus tweets in the United States just prior to the pandemic declaration. *Cyberpsychology, Behavior and Social Networking*, *23*, 889–893.

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (pp. 3844–3852).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (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* (pp. 4171–4186).

Eberts, M., & Ulges, A. (2020). Span-based joint entity and relation extraction with transformer pre-training. In *Proceedings of the 24th European Conference on Artificial Intelligence* (pp. 2006–2013).

Fu, K.-W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P., & Fung, I. C.-H. (2016). How people react to Zika virus outbreaks on Twitter? a computational content analysis. *American Journal of Infection Control*, *44*, 1700–1702.

Fu, T.-J., Li, P.-H., & Ma, W.-Y. (2019). Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1409–1418).

Glowacki, E. M., Lazard, A. J., Wilcox, G. B., Mackert, M., & Bernhardt, J. M. (2016). Identifying the public's concerns and the centers for disease control and prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. *American Journal of Infection Control*, *44*, 1709–1711.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 1025–1035).

Hang, T., Feng, J., Wu, Y., Yan, L., & Wang, Y. (2021). Joint extraction of entities and overlapping relations using source-target entity labeling. *Expert Systems with Applications*, *177*, 114853.

Hong, Y., Liu, Y., Yang, S., Zhang, K., & Hu, J. (2020). Joint extraction of entities and relations using graph convolution over pruned dependency trees. *Neurocomputing*, *411*, 302–312.

Hou, Z., Du, F., Jiang, H., Zhou, X., & Lin, L. (2020). Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: Social media surveillance in China. Preprint at *SSRN* http://dx.doi.org/10.2139/ssrn.3551338.

Hu, Y., Bai, Q., & Li, W. (2019). Context-aware influence diffusion in online social networks. In *Pacific Rim Knowledge Acquisition Workshop* (pp. 153–162). Springer.

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. ArXiv preprint arXiv:1508.01991.

Jahanbin, K., Rahmanian, V. et al. (2020). Using twitter and web news mining to predict covid-19 outbreak. *Asian Pacific Journal of Tropical Medicine*, *13*, 378.

Kassab, L., Kryshchenko, A., Lyu, H., Molitor, D., Needell, D., & Rebrova, E. (2020). On nonnegative matrix and tensor decompositions for COVID-19 Twitter dynamics. ArXiv preprint arXiv:2010.01600.

Katiyar, A., & Cardie, C. (2017). Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (pp. 917–928).

Kaveh-Yazdy, F., & Zarifzadeh, S. (2020). Track Iran's national COVID-19 response committee's major concerns using two-stage unsupervised topic modeling. *International Journal of Medical Informatics*, *145*, 104309.

Killeen, B. D., Wu, J. Y., Shah, K., Zapaishchykova, A., Nikutta, P., Tamhane, A., Chakraborty, S., Wei, J., Gao, T., & Thies, M. (2020). A county-level dataset for informing the United States' response to COVID-19. *Hospital*, *4000*, 8000.

Killgore, W. D., Cloonen, S. A., Taylor, E. C., & Dailey, N. S. (2020). Loneliness: A signature mental health concern in the era of COVID-19. *Psychiatry Research*, *290*, 113117.

Kim, E. H.-J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, *42*, 763–781.

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*.

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning* (pp. 282–289).

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 260–270).

Lamsal, R. (2021). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*, *51*, 2790–2804.

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: a text mining analysis of the centers for disease control and prevention's Ebola live Twitter chat. *American Journal of Infection Control*, *43*, 1109–1111.

Li, L., Zhang, Q., Wang, X., Zhang, J., Wang, T., Gao, T.-L., Duan, W., Tsoi, K. K.-f., & Wang, F.-Y. (2020). Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo. *IEEE Transactions on Computational Social Systems*, *7*, 556–562.

Li, W., Bai, Q., Zhang, M., & Nguyen, T. D. (2018). Automated influence maintenance in social networks: An agent-based approach. *IEEE Transactions on Knowledge and Data Engineering*, *31*, 1884–1897.

Ma, X., & Hovy, E. (2016). End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 1064–1074).

Miwa, M., & Bansal, M. (2016). End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 1105–1116).

Mohit, B. (2014). Named entity recognition. In *Natural language processing of semitic languages* (pp. 221–245). Springer.

Nelson, B. W., Pettitt, A., Flannery, J. E., & Allen, N. B. (2020). Rapid assessment of psychological and epidemiological correlates of COVID-19 concern, financial strain, and health-related behavior change in a large online sample. *PloS one*, *15*, e0241990.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1532–1543).

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 2227–2237).

Rathan, M., Hulipalled, V. R., Venugopal, K., & Patnaik, L. (2018). Consumer insight mining: aspect based Twitter opinion mining of mobile phone reviews. *Applied Soft Computing*, *68*, 765–773.

Santoso, J., Setiawan, E. I., Purwanto, C. N., Yuniarno, E. M., Hariadi, M., & Purnomo, M. H. (2021). Named entity recognition for extracting concept in ontology building on indonesian language using end-to-end bidirectional long short term memory. *Expert Systems with Applications*, *176*, 114856.

Seewald, A. K., & Kleedorfer, F. (2007). Lambda pruning: An approximation of the string subsequence kernel for practical SVM classification and redundancy clustering. *Advances in Data Analysis and Classification*, *1*, 221–239.

Shi, J., Li, W., Yang, Y., Yao, N., Bai, Q., Yongchareon, S., & Yu, J. (2021). Automated concern exploration in pandemic situations-COVID-19 as a use case. In *Pacific Rim Knowledge Acquisition Workshop* (pp. 178–185). Springer.

da Silva, J. A. T., Tsigaris, P., & Erfanmanesh, M. (2021). Publishing volumes in major databases related to COVID-19. *Scientometrics*, *126*, 831–842.

Szomszor, M., Kostkova, P., & St Louis, C. (2011). Twitter informatics: Tracking and understanding public reaction during the 2009 swine flu pandemic. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (pp. 320–323). IEEE.

Van Der Vegt, I., & Kleinberg, B. (2020). Women worry about family, men about the economy: Gender differences in emotional responses to COVID-19. In *International Conference on Social Informatics* (pp. 397–409). Springer.

Van Lent, L. G., Sungur, H., Kunneman, F. A., Van De Velde, B., & Das, E. (2017). Too far to care? measuring public attention and fear for Ebola using Twitter. *Journal of Medical Internet Research*, *19*, e7219.

Wahbeh, A., Nasralah, T., Al-Ramahi, M., & El-Gayar, O. (2020). Mining physicians' opinions on social media to obtain insights into COVID-19: Mixed methods analysis. *Journal of Medical Internet Research*, *6*, e19276.

Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. *IEEE Access*, *8*, 138162–138169.

Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. *Journal of Medical Internet Research*, *22*, e20550.

Yin, H., Yang, S., & Li, J. (2020). Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. In *International Conference on Advanced Data Mining and Applications* (pp. 610–623). Springer.

Yu, B., Zhang, Z., Shu, X., Wang, Y., Liu, T., Wang, B., & Li, S. (2020). Joint extraction of entities and relations based on a novel decomposition strategy. In *Proceedings of the 24th European Conference on Artificial Intelligence* (pp. 2282–2289).

Zeng, D., Liu, K., Lai, S., Zhou, G., & Zhao, J. (2014). Relation classification via convolutional deep neural network. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 2335–2344).

Zeng, D., Zhang, H., & Liu, Q. (2020). Copymtl: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence* 05 (pp. 9507–9514).

Zeng, X., Zeng, D., He, S., Liu, K., & Zhao, J. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. 506–514).

Zhang, Y., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 2205–2215).

Zheng, S., Wang, F., Bao, H., Hao, Y., Zhou, P., & Xu, B. (2017). Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (pp. 1227–1236).

Zhou, G., & Su, J. (2002). Named entity recognition using an HMM-based chunk tagger. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 473–480).

Zhou, G.-D., & Zhu, Q.-M. (2011). Kernel-based semantic relation detection and classification via enriched parse tree structure. *Journal of Computer Science and Technology*, *26*, 45–56.