applied sciences

MDPI

Article

Event Monitoring and Intelligence Gathering Using Twitter Based Real-Time Event Summarization and Pre-Trained Model Techniques

Chung-Hong Lee [1,*], Hsin-Chang Yang [2], Yenming J. Chen [3] and Yung-Lin Chuang [1]

[1] Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807618, Taiwan; assh70057@gmail.com
[2] Department of Information Management, National University of Kaohsiung, Kaohsiung 811726, Taiwan; yanghc@nuk.edu.tw
[3] Management School, National Kaohsiung University of Science and Technology, Kaohsiung 824005, Taiwan; yjjchen@nkust.edu.tw
* Correspondence: leechung@mail.ee.nkust.edu.tw

Abstract: Recently, an emerging application field through Twitter messages and algorithmic computation to detect real-time world events has become a new paradigm in the field of data science applications. During a high-impact event, people may want to know the latest information about the development of the event because they want to better understand the situation and possible trends of the event for making decisions. However, often in emergencies, the government or enterprises are usually unable to notify people in time for early warning and avoiding risks. A sensible solution is to integrate real-time event monitoring and intelligence gathering functions into their decision support system. Such a system can provide real-time event summaries, which are updated whenever important new events are detected. Therefore, in this work, we combine a developed Twitter-based real-time event detection algorithm with pre-trained language models for summarizing emergent events. We used an online text-stream clustering algorithm and self-adaptive method developed to gather the Twitter data for detection of emerging events. Subsequently we used the Xsum data set with a pre-trained language model, namely T5 model, to train the summarization model. The Rouge metrics were used to compare the summary performance of various models. Subsequently, we started to use the trained model to summarize the incoming Twitter data set for experimentation. In particular, in this work, we provide a real-world case study, namely the COVID-19 pandemic event, to verify the applicability of the proposed method. Finally, we conducted a survey on the example resulting summaries with human judges for quality assessment of generated summaries. From the case study and experimental results, we have demonstrated that our summarization method provides users with a feasible method to quickly understand the updates in the specific event intelligence based on the real-time summary of the event story.

Keywords: text summarization; event detection; machine learning; natural language processing; Twitter

Citation: Lee, C.-H.; Yang, H.-C.; Chen, Y.J.; Chuang, Y.-L. Event Monitoring and Intelligence Gathering Using Twitter Based Real-Time Event Summarization and Pre-Trained Model Techniques. *Appl. Sci.* **2021**, *11*, 10596. https://doi.org/10.3390/app112210596

Academic Editor: Andrea Prati

Received: 25 September 2021
Accepted: 5 November 2021
Published: 11 November 2021

1. Introduction

In emergencies with high market uncertainty and business interruption, real-time decision-making ability is usually a necessary condition to avoid risks. For example, the outbreak of COVID-19, earthquakes or floods have brought unexpected challenges to companies, making people realize that they should have real-time event information and resilience to immediately respond to changing external disasters. Furthermore, during such high-impact events, users may want to learn up-to-date information as the event develops because they want to know the possible trend of the event. In this case, it is almost impossible for anyone to browse through all of the news articles to understand the event stories required for decision-making. Automated summarization systems that access

a large number of information sources in real time can help by providing event summaries, which are updated whenever important new events are detected. The update of the event will help those interested persons to understand the situation of the current event. With the real-time nature of social media, an automated event summary system using social media (such as Twitter) data can provide people with event updates while the situation is still evolving, without the need for individuals to manually analyze a large number of news articles. The term "real-time" here means that the summary system can expand or update the summary at any point in time, and the real-time computing property is not the focus of this work. In order to meet the above-mentioned increasing demand, in this work, we provide a solution for real-time event summarization. It combines the developed Twitter-based real-time event detection algorithm with a pre-trained language model for summarizing emergency events.

Automatically summarizing social message topics is a relatively unnoticed emerging research field. Essentially, tweet summaries can be viewed as an example of a more general automatic text summarization problem, which is an issue of automatically generating a condensed version of the most important content from user-generated text content. In addition, it is also a type of stream summarization. For text streams containing millions of documents covering various topics and events, traditional single-document and multi-document summarization methods cannot solve the challenge of information overload. Rudrapal [1] summarized the methods of automatic summarization on Twitter's most popular topics. For some highly structured and recurring events in social media messages, it is best to use more complex techniques to summarize relevant information. Chakrabarti [2] formalized the summary of event messages and proposed a hidden Markov model. To learn the solution of the non-recessive state representation of the event. Chua [3] proposed a framework for searching and summarizing on social media, extracting the most representative messages from time-series social media messages to generate continuous and concise summaries. Furthermore, the text summarization algorithms can be generally divided into two types of methods, including extractive summarization algorithms and abstractive summarization algorithms. Since our task requires extensive domain knowledge beyond the meaning of individual words and sentences, in this work, we limit our focus to abstractive summarization algorithms.

In order to generate more semantically rich sentence representations, we apply the concept of transfer learning to perform text summarization with a pre-trained language model. Transfer learning, where the model is first pre-trained on data-rich tasks, and then fine-tuned on downstream tasks, has become a powerful technology in natural language processing (NLP) field. In this work, the T5 pre-trained model [4] is used to conduct experiments for text summarization. The motivation for this work can be divided into the following two parts:

- In a high-impact event, people may want to learn the latest information as the event develops. It is necessary to develop an effective real-time event summary system to obtain sufficient real-time intelligence information for effective decision-making.
- In order to acquire extensive coverage of event details, using pre-trained language models can help us expand more knowledge and trending topics about how event stories develop. The benefit of this potential for event awareness deserves in-depth study.

The view is taken, therefore, in this work, we utilized the above-mentioned approaches to produce a streamlined version of a dynamically changing event data from the collected Twitter data set; when users view this version, they will be able to immediately capture the most important part of the latest event content in real-time for making appropriate decisions. It is worth mentioning that "real-time event summarization" is different from "real-time event detection" methods. The summarization system we developed can summarize the continuously accumulated hot topic tweets into a short sentence, instead of just performing real-time event detection. The summary generated by our system can represent a simplified version of the event tweet message of interest, starting at any point in time of the ongoing event.

The rest of the paper is organized as follows. Section 2 surveys several related work in the literature. The proposed system model and system framework are described in Section 3. Section 4 reports on our pre-trained language model for text summarization. In particular, we introduce our developed text summarization method using a hybrid pre-trained model system. In Section 5, we describe our experiments in evaluation on (i) the performance of the text summarization method based on Rouge evaluation method, and (ii) the effectiveness of the real-time summary method developed for intelligence collection and event monitoring. In particular, in this work, we provide a real-world case study, namely the COVID-19 pandemic event, to verify the applicability of the proposed method. In Section 6, the discussion of experimental results is addressed in detail. Section 7 concludes the work by summarizing our research contributions.

2. Related Work

In this work, we attempt to build a real-time event summarization system which can generate a simplified version of a dynamically changing event data from the collected Twitter data set. When users read the generated sentences, they will be able to immediately capture the most important part of the latest event content in real time to make appropriate decisions. Using Twitter for event detection and event summarization is a vast subject. Twitter is also employed to better understand how the communication transfers during citizen movements and events. In our previous work [5,6] a real-time event detection approach has been well developed to gather tweets of detected events for awareness of emerging events in the real world.

Summarization of a single text summarization is the task of automatically generating a shorter version of a document and retaining its most important information. The types of automatic text summarization can be classified as extractive and abstractive. The extractive text summarization extracts sentences from the texts, while the abstractive ones may produce sentences and phrases which do not appear in the original documents. In practice, the techniques of abstractive text summarization emulate human summarization in that it utilizes a vocabulary beyond the specified text and summarizes key points in the text. Due to such approaches emulate how people summarize material, they normally require more computing power such as several GPUs to train the model over many days for learning algorithms, thus it is generally more difficult to acquire a cost-effective solution. In this work, we concentrate on abstractive summarization techniques.

While text summarization has been investigated for years, automatic summarization on social messages (e.g., tweets) is still in its infancy. Sharifi et al., used the Phrase Reinforcement algorithm to deal with tweet summarization [7]. Inouye developed a Hybrid TF-IDF algorithm to produce multiple post summaries [8].

In order to make our model to summarize text accurately, it needs to learn to "understand" the input data. Google's BERT (Bidirectional Encoder Representations from Transformers) is the perfect tool for natural language understanding. Before the emergence of the BERT model, several important techniques related to the modeling method we adopted were developed. Recently, the Recurrent Neural Networks (RNNs; including Long Short Term Memory Networks) have been widely used for many natural language processing applications. However, it needs massive amounts of data, costly data sources, and several hours of training to achieve satisfactory results. Due to the drawback of suffering from poor performance with very long sequences of data, Vaswani proposed a model called the "Transformer", replacing the RNNs and Convolutional Neural Networks (CNNs) by utilizing an architecture comprised of feed forward networks and attention mechanisms [9]. An unsupervised learning architecture called BERT (Bidirectional Encoder Representations from Transformers) has been developed and outperformed almost all existing models in the competitions for various types of NLP tasks [10]. In addition, they published several pre-trained models which can be utilized for transfer learning for many tasks and applications [10].

Text summarization is a technology that extracts the important text or phrase information from the original text and then presents a short message to the user. According to the content and method of abstracting, it can be divided into Extractive and Abstractive. Extractive summarization is a technique that uses statistics that extracts the most important words and phrases in a text according to the given weight, and then arranges and combines these features to generate a summary of the text; abstractive summarization is a technique that uses machine learning methods to generate a summary of the text after understanding the content of the text [11–14].

Most of the extractive summarization are generated by extracting human-designed words or phrases as features in the original text [15–17], so that these extracted features are the most useful and important words and phrases in the text to achieve a more accurate final summary [18]. These features are usually rated according to their importance in the title, word frequency, word position, sentence length, similarity between sentences, number of occurrences of words, and famous words [19], and then rearranged and combined according to this rating to generate the summary. The method has been studied in various applications [20–22]. For example, Isonuma [23] proposed a framework for extractive summarization using the dataset NIKKEI financial report corpus and the New York Times annotation corpus. Nallapati [24] proposed a Recurrent Neural Network (RNN) to extract the words and phrases that are relevant to the topic. Nallapati [24] proposed SummaRuNNer, a language model based on Recurrent Neural Network (RNN), which has a different training mode from the usual models and can reduce the need for extractive feature labels during training. Gunawan [25] used sentence length, headline features, and keyword frequency to develop an unsupervised summarization model for 3 K articles. Narayan [26] summed up the concept of extracted summaries as a task for ranking sentence and presents a training method that uses reinforcement learning to optimize the reward shaping that related to the task by exploring the space of candidate summaries. Dong [27] presents a framework for extractive summarization through the neural networks and enhanced learning contexts, which did not require sentence-level extraction labels on extracted summaries and optimized the ROUGE score of model that generates the summary.

Rush [28] presented a Neural Attention Model for Abstractive Sentence Summarization, combining the chance model with the generation algorithm to produce an accurate summary, and acquires good results in the Gigaword and DUC datasets, on which Chopra [29] further improves the performance of the summary after replacing the decoder with an RNN. Song [30] proposed an LSTM-combined CNN abstractive text summarization framework (ATSDL), which extracts key phrases from the text and generates a smooth and grammatical summary by deep learning models. Paulus [31] proposes a neural network model for dealing with longer textual data, and the state-of-the-art is achieved in the experiment using 12 data sets. Chen [32] proposed a sentence-level learning method that allows abstractive summarization models to learn the hierarchical nature of words and sentences. Masum [33] used a bidirectional RNN with an LSTM as the encoding layer and an attention model as the decoding layer to create a short, fluent and comprehensible summary of a text document. Cai [34] proposed an abstractive text summarization model RC-Transformer (RCT), which adds an RNN encoder to capture the context of the sequence and adds a convolutional layer to extract important sentences and understand the semantics to generate a paragraph of extremely high quality summary.

In recent years, transfer learning methods using pre-trained models have achieved good results in most natural language processing tasks [35]. The self-supervised pre-training of the model on a large amount of unlabeled data will affect the effect of transfer learning, such as building a language model or filling in the blanks of words. Therefore, fine-tuning a small data set in the pre-trained model will be more effective. Devlin [10] proposed the BERT (Bidirectional Encoder Representations from Transformers) model in 2018, which has had a great impact on the field of natural language processing. BERT provides a pre-trained bidirectional feature for unlabeled text data in the context of all levels. Automatic text summarization is a hot research topic in the field of natural language processing.

Many studies use deep learning technology to obtain important information from text data, so that the machine can understand the text and generate a text-related summary. Therefore, many recent studies have applied pre-training models to text summarization. Liu [36] proposed a framework built using BERT. The framework can provide extractive and abstractive types of text summarization. The experimental results on three data sets (CNN/DailyMail, NYT and XSum) show that their model achieves the state-of-the-art result under the automatic and human-based evaluation protocol. Kieuvongngam [37] used the BERT model and GPT-2 model to conduct text summarization on the open research data set of COVID-19 and extracted the keywords and semantic meaning of the sentences from the original text data through the above-mentioned pre-trained model. Their system provides a concise and meaningful summary for the medical literature. Khandelwal [38] used the pre-trained Transformer language model to generate summaries from the original text data. Compared with the pre-trained Transformer encoder-decoder network, this model has significantly improved results. Zhang [39] proposed a pre-trained model HIBERT (Hierarchical Bidirectional Encoder Representations from Transformers) for pre-training on text data and unlabeled data. This model can achieve higher results on the CNN/Dailymail data set and the New York Times data set. Therefore, the experimental results show that the use of pre-trained models can make large open data sets acquire better results. Farahani [40] mentioned two approaches to deal with the task of text summarization using mT5 model and ParsBERT model and got good results on a data set named pn-summary for Persian abstractive text summarization. Ma [41] proposed a pre-trained model T-BERTSum for text summarization, which captures the key words of the topic information of social media, understands the meaning of the sentence and judges the topic of the message discussion, and then generates a high-quality section Summary. Kerui [42] uses BERT, Seq2seq and reinforcement learning to form a text summary model. Garg [43] uses T5, one of the most advanced pre-trained models, to perform a summary task on a data set with 80,000 news articles, and the results indicate that the summary generated by T5 has better quality than those generated by other models. Daiya [44] has developed a pre-trained language model ENEMAbst that can be used in extractive and abstractive summarization techniques. Both abstractive summaries and extractive summaries on the CNN/DailyMail dataset obtain the most excellent results. Fecht [45] proposed a pre-trained language model that can be used for abstractive text summarization, and combined with the use of BERT to experiment on the German summarization data set. After performing the contextual word embedding task through BERT, the model can improve the model's ability to generate high-quality German summaries. In Table 1, some important related work is shown for comparison.

In summary, in our work, social media (such as Twitter), pre-trained models, and text summaries (including extractive and abstractive methods) are the three key elements of event summarization. From the above literature review, the text summaries of specific event tweets on social media have gradually become one of the most popular summarization research topics in recent years [1–3,41,42]. However, extractive summarization cannot obtain many key sentences in the tweet-form data, resulting in unsatisfactory quality of the generated summaries. Furthermore, due to our task requires extensive domain knowledge beyond the meaning of individual words and sentences, in this work we limit our focus to abstractive summarization algorithms. On the other hand, in the recent references of text summarization, there are more and more work using pre-training language models [35–44]. Among them, BERT is the pre-trained model mainly used by most research teams. Most researchers use the improved BERT model architecture in their work [36,37,39–41]. However, because the sentences in the pre-training data set of BERT have correct grammar, BERT does not perform well for tweets with less grammar. In order to simplify the summarization process, the text-to-text pre-trained model (e.g., T5 model) has become a good choice. Unfortunately, for our problem domain, so far little attention has been paid to the development of a summary method that supports real-time functions to achieve the requirements of real-time event monitoring and intelligence collection. Motivated by this, in this work, we developed a novel model combining a real-time event summarization

system and a Twitter-based event detection method. The developed summarization system can summarize the continuously accumulated hot topic tweets into a short sentence, instead of just performing real-time event detection. To the best of our knowledge, this is the first attempt to develop such an event summarization system with multiple functional requirements to meet the challenge of intelligence gathering in emergency situations.

Table 1. Related work of automated text summarization.

Citing the Author	Model	Using Pre-Trained Model	Method of Text Summarization	Data Set
Isonuma et al. [23]	CNN-LSTM	No	Extractive	NIKKEI Financial Report
Nallapati et al. [24]	SummaRuNNer	No	Extractive	CNN/Daily Mail and DUC 2002
Masum et al. [33]	Bidirectional LSTM	No	Abstractive	Amazon Fine Food Reviews
Fecht et al. [45]	BERT	Yes	Abstractive	German Wikipedia articles
Kerui et al. [42]	BERT	Yes	Abstractive	LCSTS
Garg et al. [43]	T5	Yes	Abstractive and Extractive	80 K News Articles
Farahani et al. [40]	mT5	Yes	Abstractive	Pn-summary
Cai et al. [34]	RC-Transformer	Yes	Abstractive	Gigaword and DUC 2004
Chopra et al. [29]	Conditional Generation by RNN	No	Abstractive	
Song et al. [30]	CNN-LSTM	No	Abstractive	
Paulus et al. [31]	Attention-RNN	No	Abstractive	
Ma et al. [41]	T-BERTSum	Yes	Abstractive	CNN/Daily Mail
Zhang et al. [39]	HIBERT	Yes	Abstractive	
Liu et al. [36]	BERT	Yes	Abstractive and Extractive	
Daiya [44]	ENEMABst	Yes	Abstractive and Extractive	

3. Models and Methods

3.1. System Framework

In this work, the pre-trained models are mainly used to perform summarization function on social media data to fine-tune the text summaries. With the rapid development of social media, many messages flow and gather on social media, forming various event topics. However, identifying "event" is one of the most difficult challenges of current decision support systems. If the system encounters an emergency, it is usually unable to promptly notify users of the warning to avoid risks. A sensible solution is to employ the functions of the real-time event summarization to obtain emergent event information. As a result, the system framework of this study utilizes the method of event detection on social-media and messages to extract their main content, and then summarized to understand the evolution of important events. Figure 1 illustrates the conceptual diagram of the process followed to extract real-time event intelligence for both the event detection and summarization. The system is designed as a two-pass process, consisted of two modules, say, real-time event detection module and event-message accumulation and summarization module. More details on real-time event summarization are presented in the following. Firstly, in this work, we combine a developed Twitter-based real-time event detection algorithm with pre-trained language models for summarizing emergent events. We used an online text-stream clustering algorithm and self-adaptive method developed to gather the Twitter data for detection of major international events. The unimportant messages on the data set are removed through the trained SBERT based classifier. After that, the DBSCAN algorithm is used to cluster similar topics in each day into hot topics (i.e., event topics). These event topics in Twitter data are generated by using the algorithm we developed to generate storylines. When the hot topics of each day are summarized, the specific topic and other hot topics are semantically correlated. When the correlation exceeds a certain threshold, it is added to the storyline. After daily updates, through the

Twitter based real-time intelligence data collection, a storyline of a specific event is jointly generated.

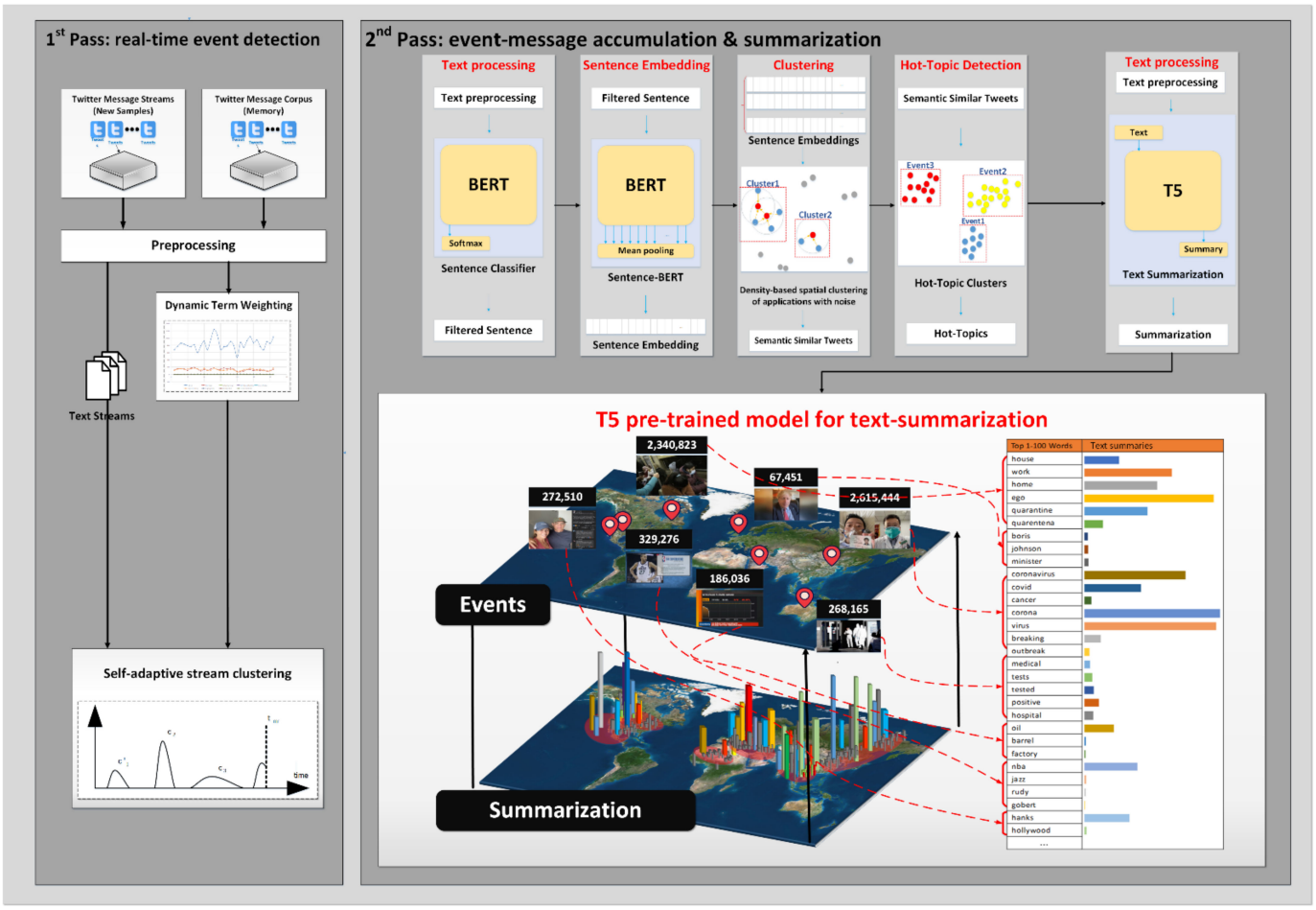


Figure 1. Proposed Twitter based real-time event summarization and pre-trained models.

The next work is to deal with text summarization of specific topics. In this work, we use the pre-trained model (i.e., T5; Text-to Text Transfer Transformer) to experiment with the Xsum data set to perform text summarization, and the Rouge metrics are used to evaluate the resulting summaries. The Xsum [46] data set contains 226,711 news articles and their summaries. News articles come from the BBC's content from 2010 to 2017. The articles cover many fields such as news, politics, sports, weather, business, technology, science, health, family, education, entertainment and art, etc. We split the data set into 90% training data set, 5% validation data set and 5% test data set. The pre-trained language models and pre-training process will be described later in more detail.

Subsequently, we started to use the trained model to summarize the incoming Twitter data set for experimentation. The large amount of events and related tweets detected by the real-time event detection system are grouped into about hundreds of event topics every day. Taking the Twitter data in 2020 as an example, there are about 36,000 event topics on Twitter. Using the resulting models of text summarization on popular topics of social media in this work, the system is used to perform real-world event summarization tasks and analyze the results. More details about pre-trained language models for text summarization are presented in the following section.

3.2. Proposed Approach for Collectively Summurizing Text Streams

The first part is the real-time event detection algorithm, which analyzes the life cycle of events and the distribution of word weights through dynamic term weighting [47] in the instant message stream. In the early stage of the event, potential major events are captured, and then the real-time event cluster is combined into an event sequence through the self-adaptive stream clustering method. The algorithm for real-time event detection is shown in Algorithm 1.

Algorithm 1 1st Pass real-time event detection

INPUT: TS{1, 2, . . . , m} // Twitter Streams
OUTPUT: RTED{1, 2, . . . , ε}; //Real-Time Event Detection
1: Begin
2: **for** each TS{i}
3: BursT ← Dynamic Term Weighting(i)
4: RTED ← Self-Adaptive Stream Clustering(BursT, i)
5: **end for**
6: End

$$BursT_{w,t} = BS_{w,t} \times TI_{w,t} \tag{1}$$

$$BS_{w_r^q,t} = max\left\{\frac{ar_{w_r^q,t} - E\left(ar_{w_r^q,t}\right)}{E\left(ar_{w_r^q,t}\right)}, 0\right\} \tag{2}$$

$$TI_{w_r^q,t} = 1 - \frac{-1}{ln\left|M^{[t_g,\ t]}\right|} * ln\frac{\left|\ w_r^q\right|}{\left|m:\ w_r^q \in W^{[t_l,\ t]}\right|} \tag{3}$$

Dynamic Term Weighting can be divided into two factor values: Burst Score, *BS* and Term Importance, *TI*, as shown in Formula (1). In Formula (2), Burst Score is to calculate the burstiness of a keyword according to the rate at which the keyword arrives. In Formula (3), the key word weight is the logarithm of the ratio of the number of documents in the global time window to the number of messages containing the keyword to determine the popularity of the keyword [47].

Self-adaptive stream clustering algorithm [48] is used to represent the situation that, when a new message flows into the system, the message will be compared with the existing online group for similarity. When the similarity comparison result is greater than the threshold, the new message will be added to the most similar group in the online group. If it is less than the threshold, the system will create a new group and add the message. At the same time, the system will remove non-informative messages and aging online groups based on the adjustable sliding window module and the time decay function. The proposed self-adaptive stream clustering algorithm is shown in Algorithm 2.

Once the adaptive stream clustering algorithm is used to process stream messages in real time, the number and weight of features will be automatically adjusted within the thread time width tw. As shown in the Algorithm 2, when a new message flows into the system, the message will be compared with the existing online cluster for similarity evaluation. When the similarity comparison result is greater than the threshold, the new message will be added to the most similar cluster. If it is less than the threshold, the system will create a new cluster and add the message. At the same time, the system will remove non-informative messages and aging online clusters based on the adjustable sliding window.

Algorithm 2 Self-adaptive stream clustering

INPUT: MS{1, 2, . . . , m} // Message Streams
OUTPUT: RTEC{1, 2, . . . , ε}; HEC{1, 2, . . . , n} //Real-Time Event Cluster
1: Begin
2: **for** each MS{i}
3: **for** each RTEC{j}
4: **if**(d(·) – Θ < 0) **then**
5: add RTEC{j} into HEC and revom RTEC{j};
6: **else**
7: simOfMS&RTEC ← sim(MS{i}, RTEC{j})*d(·);
8: **if**(simOfMS&RTEC > maxSimOfMS&RTEC) **then**
9: maxSimOfMS&RTEC ← simOfMS&RTEC;
10: idOfMaxC ← j;
11: **else** continue;
12: **end for**
13: **if**(maxSimOfMS&RTEC < Θ) **then**
14: create a new RTEC for MS{i};
15: **else**
16: merge MS{i} to RTEC{idOfMaxC};
17: **end for**
18: until no messages is posted

The second part of the event-message accumulation and summarization algorithm is to filter the event information through the BERT classifier, and then use the vector of the filtered message obtained by the Sentence-BERT model and put it into the DBSCAN algorithm to cluster the event information into For each sub-event, the summary content is finally obtained through the T5 model. The proposed method for event-message accumulation and summarization algorithm is shown in Algorithm 3.

Algorithm 3 2nd Pass event-message accumulation & summarization

INPUT: TM{1, 2, . . . , n} // Text Message
OUTPUT: HT{1, 2, . . . , ε}; HTS{1, 2, . . . , m}; EHT{1, 2, . . . , k} // Hot Topic
1: Begin
2: **for** each TM{i}
3: Text ← BERT Classifier Model(TM{i})
4: Sentence Embedding ← Sentence-BERT Model(Text)
5: HT ← DBSCAN(Sentence Embedding)
6: **for** each HT{j}
7: HTS ← T5 Text Summarization Model(j)
8: **end for**
9: **end for**
10: End

4. Proposed Pre-Trained Language Models for Text Summarization

In this study, the pre-trained language model is applied to social media data to fine-tune the text summaries. With the development of the Internet generation, people prefer to express their feelings or the content they are interested in on social media, and many messages flow on social media every day. When there is more and more information on a specific topic, it is easy to form a hot topic. This research will focus on summarizing popular topics when social media content forms as the main core content, and then summarize them to explore the evolution of popular topics.

Transfer learning, where the model is first pre-trained on data-rich tasks, and then fine-tuned on downstream tasks, has become a powerful technology in natural language processing (NLP) field. In this work, the T5 [4] pre-trained model is used to conduct experiments and the Rouge metrics are used to evaluate the resulting summaries. In addition, we also use the BERT pre-training model for experiments as a baseline for

performance comparison with the T5 model. The concepts of T5 and BERT pre-trained models are briefly described in the following subsections.

4.1. Bert Model and Sentence-BERT Model

The establishment of the BERT model uses other techniques for training context-dependent representation models, including Semi-supervised Sequence Learning, GPT [49], ELMo [50] and ULMFit [51], but these models are unidirectional or shallow Bidirectional, which makes these models only use the words on their left or right to learn context, while BERT uses a deep bidirectional architecture so that the model can learn the words on its left and right. The BERT model also has good results in the task of semantic similarity, but when the BERT model calculates the semantic similarity, it needs to put two sentences into the model at the same time and calculate the information of the two sentences. This approach requires a lot of For calculation and time, if there are N sentences, it needs to calculate (N∗(N + 1))/2 times, and the similarity retrieval of 10,000 sentences requires about 50 million calculations, and it takes about 50 million times to execute using the BERT model. 65 h, which makes the BERT model unsuitable for semantic similarity retrieval and cluster analysis.

Reimers proposed a model Sentence-BERT (SBERT) to improve the BERT architecture, using the twin and triplets network architecture to generate a sentence embedding that can compare semantics with cosine similarity, making it possible to calculate the similarity of 10,000 sentences The time required for the BERT is shortened from 65 h to 5 s, and the accuracy of the similarity calculation using the BERT model can still be maintained. SBERT added a layer of pooling operation behind the output layer of BERT to generate a revised sentence embedding. The maximum value is used as the sentence vector of the entire sentence. In Reimers' experiment, the Mean-Pooling method is used as the sentence vector to acquire the best results.

4.2. T5 (Text-to-Text Transfer Transformer) Model

T5 is a transformer-based model from Google that is trained in an end-to-end manner with text as input and modified text as output. It achieves state-of-the-art results on multiple NLP tasks such as summarization, question answering, machine translation using a text-to-text transformer trained on a large text corpus. The T5 model is composed of the encoder-decoder scheme of transformer model. The encoder and decoder on the T5-Small model are composed of 6-layer Transformer, the T5-Base model is composed of the BERT-base encoder and decoder, and the T5-Large model is composed of the BERT-large encoder and decoder group. T5-3B and T5-11B each increase the size of the model by 4 times. One of the goals of pre-training process is to provide general knowledge for the model to improve its performance on downstream tasks. T5 uses Transformer's Encoder-Decoder model during pre-training process. In the transfer learning of natural language processing, previous studies have used large unlabeled data sets for unsupervised learning. T5 attempts to measure the impact of the quality, characteristics and size of these unlabeled data sets, using C4 (Colosal Clean Crawled Corpus) corpus is used as the training data.

4.3. Developed Text Summarization Method Using a Hybrid Pre-Trained Model System

Although there are a lot of content from social media platforms, there are often some unimportant daily message content. Therefore, when data is processed, tweets are also filtered through classification models to filter out unimportant daily tweets. We use about 200,000 HuffPost news headline data sets from 2012 to 2018 [52]. This data set contains 40 different categories of news headline categories, plus the unimportant daily tweet content from our collected tweets and mark it as the 41st category 'OTHERS'. The text data are classified into 41 categories in total. The classification data is divided into three parts, 60% of the data is used to train the model, 20% of the data is used to verify whether the model has learned useful information, and 20% of the data is used to test the accuracy of the model. Since there are about 40,000 to 500,000 non-retweeted tweets in the daily collected

tweets, which contain a large number of daily trivial tweets, such as happy birthday, and holiday wishes, in this work the messages are filtered to avoid too much unnecessary data and reduce the calculation time. In order to alleviate this problem, we decided to use the characteristics of the BERT model as a classier to perform the function of tweet filtering. Subsequently, the filtered data is clustered through an unsupervised learning algorithm, and the content of the tweet is retained by the density of the content of the tweet, so as to ensure that the tweet is a hot topic on the day.

As a result, in this system, the tweets are aggregated into a daily form and screened and filtered the content of tweets that were not reposted in English, and filtered the tweets through the BERT classification method, using the HuffPost data set plus the tweet content from the tweets. The nonessential daily tweets are filtered out and divided into 41 categories in total. The filtered tweets are processed through the SBERT (Sentence-BERT) model to obtain sentence vectors, and the Euclidean distance between the vectors is used to calculate the similarity between sentences. Here we use the DBSCAN algorithm to find more intensive discussion content, and finally form a daily hot topic. This study will learn text summaries by fine-tuning the pre-trained model in the Xsum data sets. In the experiment, the T5 pre-trained model is firstly used for text summarization, as shown in Figure 2. Subsequently, we compare the T5 model with the BERT pre-trained model for summarization performance comparison.

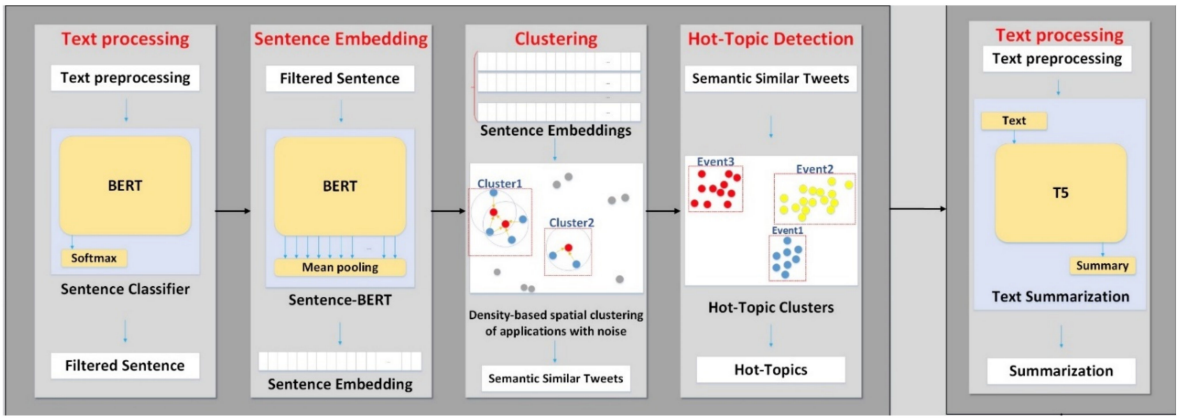


Figure 2. Developed text summarization method using a hybrid pre-trained model system.

4.4. Evaluation Method for Text Summarization

This research work mainly uses Rouge (Recall-Oriented Understudy for Gisting Evaluation) [53] as the evaluation method of text summarization. It uses automatically generated summaries and a set of reference summaries to do calculations to measure the similarity between automatically generated summaries and reference summaries. *Rouge-N* is used to calculate the number of n-grams shared by the generated summary and the reference summary, as shown in Formula (4).

$$Rouge - N = \frac{\sum_{S \in \{ReferemceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferemceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (4)$$

Rouge-L is used to match the longest common subsequence (Longest Common Subsequence, LCS) between two texts.

5. Experiment and Case Study

We designed two sets of experiments to evaluate (i) the performance of the text summarization method based on Rouge evaluation method, and (ii) the effectiveness of the real-time summary method developed for intelligence collection and event monitoring, through the implementation of a case study in the real world.

5.1. Experiment and Performance Comparison

We first performed experiments on the Xsum dataset on the BERT and T5 pre-trained models. After the evaluation of the text summaries, it can be found that the performance and output length of different pre-training models used in Tables 2–4 are different. Compared with BERT, T5 is more excellent in the performance of text summarization. The training time spent is much less than that of BERT, so the T5-base pre-trained model and the output length of 150 parameters are finally used in the text summary tool. Table 5 shows the result with the complete training data and most suitable parameters in the experiment. In order to save the training cost of the pre-trained model, the training data is compressed to find the most suitable parameters. Table 6 illustrates a comparison with other models.

Table 2. Performance evaluation (output length: 80).

Model/Score	Rouge-1	Rouge-2	Rouge-L
T5-small	28.17	8.08	21.76
T5-base	33.01	11.46	25.76
BERT	23.50	6.51	17.62

Table 3. Performance evaluation (output length: 100).

Model/Score	Rouge-1	Rouge-2	Rouge-L
T5-small	28.21	8.37	21.66
T5-base	33.64	11.80	26.34
BERT	23.42	6.56	17.43

Table 4. Performance evaluation (output length: 150).

Model/Score	Rouge-1	Rouge-2	Rouge-L
T5-small	28.97	8.26	22.07
T5-base	33.70	12.06	26.47
BERT	23.66	6.82	17.75

Table 5. Performance evaluation (using complete training data).

Model/Score	Rouge-1	Rouge-2	Rouge-L
T5-small	32.64	11.37	25.66
T5-base	38.13	16.39	30.88
BERT	29.58	11.14	22.22

Table 6. Performance evaluation (model comparison).

Model	Dataset	Rouge Score
T5-base	Xsum	38.13
UniLm	CNN/DailyMail	43.08
BERT	CNN/DailyMail	44.16

5.2. Xsum and Tweet Data Sets for Fine-Turning and Downstream Task

In this study, we use the Xsum data set and a pre-trained language model, the T5 model, to train the summary model. Since the downstream task is to generate a summary of the tweet data set, the features required for the summary generated in the proposed system are mainly learned by fine-tuning the pre-trained T5 model in the Xsum data set. The Xsum dataset contains 226,711 BBC news articles and their summaries. Articles cover news, politics, sports, weather, business, technology, science, health, family, education, entertainment, art and other fields. The coverage and diversity of Xsum are very suitable for our downstream task, which is to summarize a large number of tweets. Therefore,

after the fine-tuning process, the well-trained model is used in massive tweet data sets for summarization.

5.3. Detected Hot Topics with Twitter Data

In the hot topic detection system, a large amount of social media content is integrated into about hundreds of hot topics every day. Taking the Twitter data in 2020 as an example, there are about 36,000 hot topics in total, as shown in Figure 3. Based on the method of text summarization on popular topics in social media, the technique is applied to the task of text summarization in various fields on social media and specific data analysis.

20200101 max51	20200201 max79	20200301 max69	20200401 max103	20200501 max122	20200601 max104
20200102 max59	20200202 max48	20200302 max69	20200402 max94	20200502 max80	20200602 max90
20200103 max73	20200203 max80	20200303 max61	20200403 max105	20200503 max88	20200603 max99
20200104 max61	20200204 max75	20200304 max81	20200404 max105	20200504 max99	20200604 max123
20200105 max67	20200205 max73	20200305 max66	20200405 max95	20200505 max111	20200605 max91
20200106_max77	20200206_max80	20200306_max68	20200406_max85	20200506_max96	20200606_max100
20200107_max76	20200207_max76	20200307_max81	20200407_max106	20200507_max91	20200607_max99
20200108 max69	20200208 max67	20200308 max72	20200408 max90	20200508 max90	20200608 max91
20200109 max77	20200209 max72	20200309 max75	20200409 max83	20200509 max91	20200609 max102
20200110 max89	20200210 max75	20200310 max73	20200410 max121	20200510 max97	20200610 max102
20200111 max70	20200211 max73	20200311 max74	20200411 max101	20200511 max96	20200611 max107
20200112 max75	20200212 max69	20200312 max90	20200412 max103	20200512 max100	20200612 max103
20200113 max70	20200213 max67	20200313 max71	20200413 max90	20200513 max89	20200613 max99
20200114 max61	20200214 max65	20200314 max71	20200414 max101	20200514 max125	20200614 max104
20200115 max62	20200215 max69	20200315 max82	20200415 max103	20200515 max99	20200615 max103
20200116 max78	20200216 max73	20200316 max83	20200416 max82	20200516 max91	20200616 max94
20200117_max66	20200217_max63	20200317_max89	20200417_max96	20200517_max93	20200617_max101
20200118_max60	20200218_max81	20200318_max97	20200418_max140	20200518_max97	20200618_max112
20200119 max84	20200219 max75	20200319 max74	20200419 max131	20200519 max92	20200619 max97
20200120 max77	20200220 max82	20200320 max141	20200420 max80	20200520 max90	20200620 max86
20200121 max72	20200221 max76	20200321 max80	20200421 max104	20200521 max89	20200621 max99
20200122 max77	20200222 max65	20200322 max77	20200422 max106	20200522 max108	20200622 max103
20200123 max72	20200223 max71	20200323 max88	20200423 max114	20200523 max95	20200623 max100
20200124 max86	20200224 max74	20200324 max95	20200424 max105	20200524 max74	20200624 max103
20200125 max78	20200225 max78	20200325 max85	20200425 max96	20200525 max80	20200625 max95
20200126 max71	20200226 max76	20200326 max108	20200426 max82	20200526 max92	20200626 max121
20200127_max71	20200227_max78	20200327_max104	20200427_max95	20200527_max96	20200627_max103
20200128_max75	20200228_max83	20200328_max92	20200428_max94	20200528_max109	20200628_max112
20200129 max81	20200229 max63	20200329 max99	20200429 max91	20200529 max101	20200629 max97
20200130 max77		20200330 max94	20200430 max97	20200530 max95	20200630 max108
20200131 max70		20200331 max118		20200531 max86	

20200701 max107	20200801 max81	20200901 max130	20201001 max139	20201101 max127	20201201 max123
20200702 max95	20200802 max108	20200902 max107	20201002 max146	20201102 max115	20201202 max129
20200703 max114	20200803 max92	20200903 max117	20201003 max100	20201103 max118	20201203 max137
20200704 max85	20200804 max105	20200904 max105	20201004 max101	20201104 max127	20201204 max122
20200705 max83	20200805 max90	20200905 max88	20201005 max121	20201105 max132	20201205 max114
20200706_max101	20200806_max104	20200906_max67	20201006_max111	20201106_max126	20201206_max125
20200707_max108	20200807_max139	20200907_max96	20201007_max101	20201107_max111	20201207_max108
20200708 max98	20200808 max112	20200908 max105	20201008 max101	20201108 max141	20201208 max110
20200709 max93	20200809 max84	20200909 max104	20201009 max108	20201109 max105	20201209 max135
20200710 max115	20200810 max95	20200910 max115	20201010 max126	20201110 max118	20201210 max127
20200711 max107	20200811 max110	20200911 max105	20201011 max93	20201111 max132	20201211 max136
20200712 max100	20200812 max106	20200912 max102	20201012 max82	20201112 max118	20201212 max127
20200713 max94	20200813 max112	20200913 max113	20201013 max104	20201113 max120	20201213 max128
20200714 max93	20200814 max113	20200914 max102	20201014 max106	20201114 max107	20201214 max114
20200715 max96	20200815 max122	20200915 max108	20201015 max145	20201115 max108	20201215 max129
20200716 max92	20200816 max94	20200916 max119	20201016 max100	20201116 max113	20201216 max110
20200717_max100	20200817_max83	20200917_max131	20201017_max118	20201117_max124	20201217_max121
20200718_max86	20200818_max99	20200918_max103	20201018_max121	20201118_max121	20201218_max131
20200719 max98	20200819 max94	20200919 max105	20201019 max113	20201119 max135	20201219 max111
20200720 max107	20200820 max115	20200920 max97	20201020 max93	20201120 max163	20201220 max97
20200721 max96	20200821 max115	20200921 max118	20201021 max127	20201121 max111	20201221 max121
20200722 max101	20200822 max88	20200922 max101	20201022 max128	20201122 max112	20201222 max123
20200723 max134	20200823 max92	20200923 max125	20201023 max160	20201123 max117	20201223 max105
20200724 max139	20200824 max112	20200924 max98	20201024 max122	20201124 max123	20201224 max100
20200725 max111	20200825 max113	20200925 max97	20201025 max134	20201125 max141	20201225 max97
20200726 max105	20200826 max111	20200926 max111	20201026 max134	20201126 max122	20201226 max93
20200727_max103	20200827_max113	20200927_max109	20201027_max110	20201127_max122	20201227_max95
20200728_max105	20200828_max116	20200928_max118	20201028_max153	20201128_max107	20201228_max114
20200729 max111	20200829 max104	20200929 max105	20201029 max130	20201129 max125	20201229 max91
20200730 max102	20200830 max87	20200930 max142	20201030 max128	20201130 max101	20201230 max107
20200731 max110	20200831 max116		20201031 max94		20201231 max105

Figure 3. Daily detected event hot topics in 2020.

5.4. Case Study:Tweet Summary of the "COVID-19 Pandemic" Event

In 2020, COVID-19 is undoubtedly one of the hottest topics. The COVID-19 pandemic has exposed many serious errors in global public policy preparation and response. Although it has been affected by other epidemics and pandemics for more than a decade, many countries have failed to incorporate lessons learned into their pandemic preparedness and response plans. One of the main factors is that we lack effective real-time event-related intelligence collection systems which can provide a summary of updates related to the COVID-19 pandemic in all countries. Hence, we cannot learn from the successful experience of other countries in fighting the epidemic in time. Since the COVID-19 outbreak, very few studies that respond to practical needs, such as through rapid advice guidelines or application modules, have been conducted. However, many professionals, such as health workers, are in urgent need of effective ways to help people improve their well-being during this pandemic. According to the US Centers for Disease Control and Prevention (CDC), information on the outbreak of the coronavirus disease may increase the levels of distress among specific population—especially in those who were already struggling with mental health issues before the pandemic. They may be likely to experience higher levels of worry and stress regarding issues surrounding COVID-19. This trepidation may worsen over time and cause a long-lasting mental disturbance [54]. We wish that the proposed system can help the professionals detect and track COVID-19 events at an early stage as well as monitoring the emergency events within specific groups.

The view is taken, therefore, in this case study, we combine our developed Twitter-based real-time event detection algorithm with pre-trained language models for summarizing COVID-19 events and sub-events.

In Figure 4, it can be found that there are many detected hot-event topics around COVID-19 in 2020. In order to satisfy users' information needs, we pick up all of the real-time event topics with a comprehensive spatio-temporal viewpoint. By analyzing the events detected by our system, the spatial and temporal impacts of the emerging events can be estimated. Therefore, we first build a representation of the spatial distribution by mapping the summary of events related to COVID-19 in the world in 2020, as shown in Figure 5. After that, in terms of temporal aspects, we constructed a storyline retrospection mechanism that demonstrates the story telling capability for users to understand the context of event development in a timeline, as illustrated in Figures 6 and A1–A3.

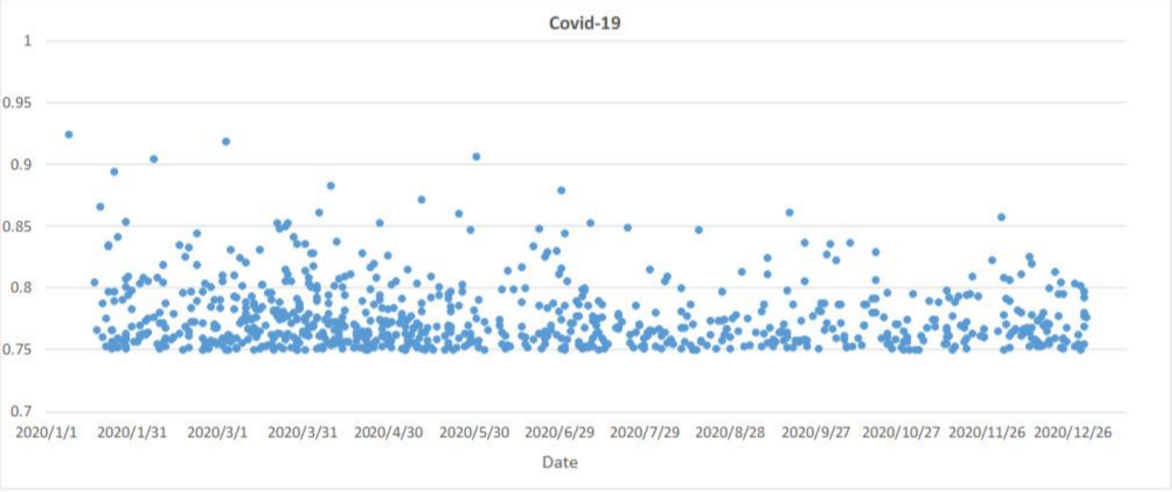


Figure 4. Detected event topics related to COVID-19 in 2020.

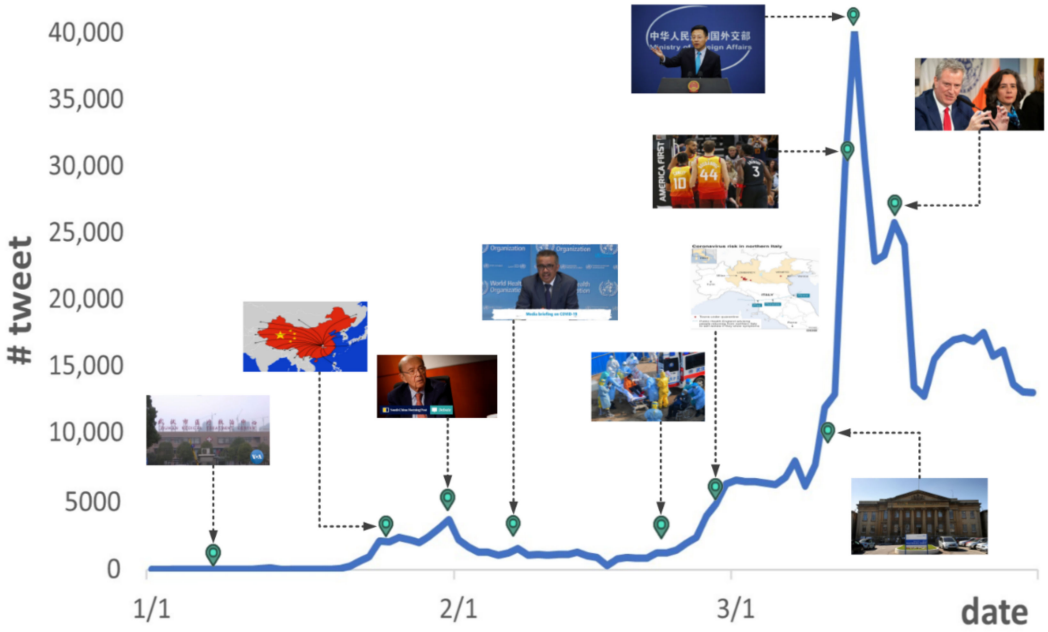
Figure 5. Representation of a map of generated summaries related to COVID-19 events in 2020.

In Figure 5, we present hot topics in the form of a map to illustrate global events to understand the spatial impact of the COVID-19 pandemic on the world.

Furthermore, in this work, we developed a storyline retrospection mechanism that demonstrates the story telling capability for users to understand the context of event development in an efficient way. It can be used to identify events and their relations under an event topic and then composes a storyline which gives users the sketch of event evolution in the topic, by computing daily related tweets and summaries of formulated event topics. The main storyline construction removes irrelevant events and presents a main theme that exhibits the sequence of events in a hot topic.

In Figures 6 and A1–A3, we can see the storyline of COVID-19 in 2020, from the beginning of the outbreak to the emergence of the final vaccine.

Figure 7 shows the example resulting summaries of COVID-19 event related tweets in 2020, starting with a new virus related to the SARS outbreak that may cause Chinese health officials to discuss on social media Pneumonia. From the time trend shown in Figures A1–A3, when the COVID-19 pandemic first broke out, social media were mainly discussing where the pandemic occurred and how to properly protect themselves from the pandemic. After a period of time, the social media began to gradually appear on the topic of what drugs may be able to treat the symptoms of COVID-19 or the topic of vaccine-related research and development. As time progressed, social media began to see related topics such as whether the vaccine was safe.



2020/1/9	A new virus linked to the SARS outbreak may be causing pneumonia in China, health officials say.
2020/1/18	Health officials in the US are to begin screening passengers from China for a new virus that has killed two people.
2020/1/19	A new virus has been discovered in the Chinese city of Wuhan, prompting US officials to screen passengers from the region.
2020/1/20	Shares in China have risen sharply as new cases of a mysterious coronavirus surge.
2020/1/21	A new strain of coronavirus has been confirmed in China, extending the reach of the disease to humans.
2020/1/23	The death toll from a new coronavirus outbreak in China has risen to at least nine, as the world's second largest city is locked down.
2020/1/24	The world's second largest city is on lockdown as a new Coronavirus outbreak threatens to kill millions of people.
2020/1/25	A new coronavirus outbreak in China has spread to several countries, with the death toll now at 67.
2020/1/28	The Centers for Disease Control (CDC) is monitoring possible coronavirus cases across US states.
2020/1/29	The US government has warned that the Coronavirus outbreak in China will have a "significant impact" on the Chinese economy.
2020/1/30	The Coronavirus is a virus that's caused by coronavirus.
2020/1/31	US Commerce Secretary Wilbur Ross says a Coronavirus outbreak in China will help to boost US jobs.
2020/2/1	The latest outbreak of the Coronavirus.
2020/2/2	The Coronavirus outbreak in the US state of California is causing panic, according to the World Health Organization.
2020/2/3	A man in the Philippines has died from a Coronavirus (nCoV) infection that has spread beyond mainland China for the first time.
2020/2/4	The death toll from the Novel Coronavirus nCoV virus in China has reached a record high, according to official figures.
2020/2/5	The World Health Organization (WHO) is urging member countries to share data on the coronavirus.
2020/2/6	The World Health Organization (WHO) has asked m in funding to fight the novel Coronavirus.
⋮	⋮
2020/3/13	The Corona virus originated in China, according to the Chinese government.
2020/3/14	A fake coronavirus map is being used to spread malware via email.
2020/3/15	The US government has said it is preparing for a coronavirus pandemic that could kill up to half of the population by 2050.
2020/3/16	US President Donald Trump has vowed to shut down the economy and close schools in New York City because of a coronavirus outbreak.
2020/3/17	The US is shutting down all but essential government services as a result of the Coronavirus pandemic.
2020/3/18	New York City is considering a shelter in place order amid an outbreak of coronavirus.
2020/3/19	The US is facing a major health crisis after the coronavirus outbreak, which has killed more than 11,000 people.
2020/3/20	US President Donald Trump has called the Coronavirus a hoax and warned that it could cost millions of dollars in damages.
2020/3/21	Donald Trump has said a virus that originated in China came from China.
2020/3/22	The Corona virus was made in China, according to a report by the BBC.
2020/3/23	The US Senate has voted to allow the National Guard to carry out life-saving health checks for people who have been infected with the Zion virus.
2020/3/24	Cyber-criminals are using fake Coronavirus to spread malware, according to a report by the BBC.
2020/3/25	President Donald Trump has vowed to cut spending on the coronavirus virus, which could kill up to tens of thousands of Americans.
2020/3/26	US President Donald Trump is facing a "pandemic" of the Corona virus, which will kill millions of Americans.
2020/3/28	Apple and the White House have released new apps to help protect against the spread of the Coronavirus virus.
2020/3/29	The death toll from the Coronavirus outbreak in the US has risen to more than 100, according to US officials.
2020/3/30	The UK is facing a six-month lockdown on the spread of Coronavirus after it was warned restrictions could last for up to six months.
2020/3/31	The World Health Organization is working to help the world's most vulnerable people and those living in refugee camps, who are affected by a pandemic.

Figure 6. Generated storyline related to COVID-19 from January to March 2020.

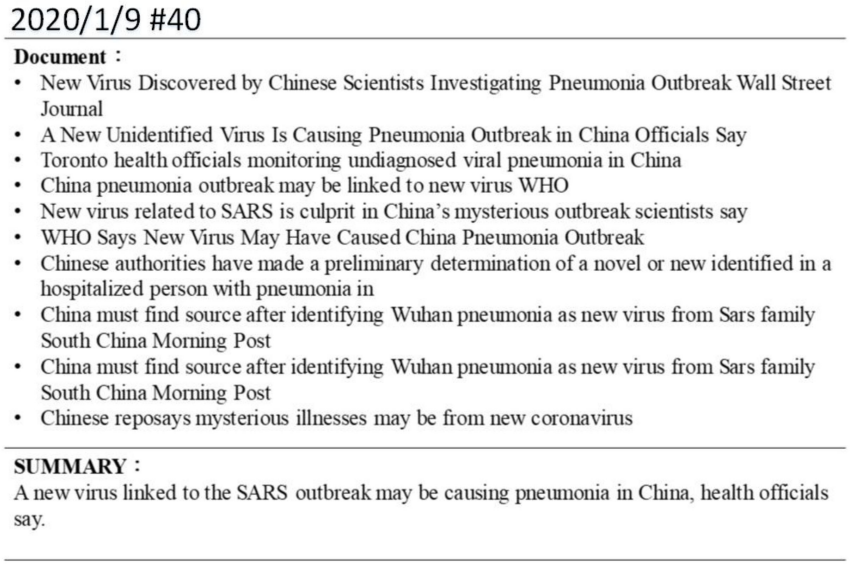


Figure 7. A sample resulting summary of formulated event topics (9 January 2020).

5.5. Human Evaluation for Quality Assessment of Generated Summaries

In order to verify the quality of the text summary results, we conducted a survey on the example resulting summaries (in Figures 7 and A4–A11) with human judges for quality assessment of generated summaries. In this work, we invited 39 participants from University graduate students, including 19 males and 20 females. Their age ranges from 20 to 30 years old. Participants were asked to rate the summaries on a 1–5 scale, with higher being better. The assessed summaries were selected using random sampling. Thus, we randomly selected 10 sample summaries from our summarization system for evaluation. In this work, the summaries are evaluated from three aspects. The first aspect is to assess the consistency of the topic between the tweets and the summary. The second aspect is to evaluate the closeness to the "facts" of the summary. The third aspect is to assess the degree of grammatical correctness of the summary. The summaries are reviewed on a scale of 1 to 5 points, and the evaluation results are shown in Table 7. Options for grading summaries are as follows: 1: Bad 2: Below Average 3: Average 4: Good 5: Excellent. In Table 7, we find that the mean values (i.e., 3.92, 3.89, 3.92, respectively) are close to 4 points, which show that the experimental result is acceptable and very close to the performance of human. In addition, we used Cronbach's alpha test to check whether the multi-question Likert scale survey is reliable and acquires results (Cronbach's alpha coefficients are 0.725, 0.812, 0.846). It shows that Cronbach's alpha coefficient is greater than 0.7, and the resulting coefficient indicates that the results are reliable.

Table 7. Quality assessment of generated summaries.

	Consistency of the Topic	Closeness to the Facts	Grammatical Correctness
No. of Participant	39	39	39
Minimum	2	2	1
Maximum	5	5	5
Mean	3.92	3.89	3.92
Median	4	4	4
Std. Deviation	0.824	0.819	0.928
Std. Error	0.132	0.131	0.148
Cronbach's alpha	0.725	0.812	0.846

6. Discussion

In this work we use natural language processing techniques to deal with social media data, extract event topics from daily social messages, and summarize them to quickly grasp various event intelligence information. In the experiment, we adopted a pre-trained model (i.e., T5) method to quickly fine-tune the model to learn text summaries through the Xsum data set, which was achieved a score of 38.13 in the evaluation of Rouge-1. After that, the trained model is actually applied to real social media data. We started to use the trained model to summarize the incoming Twitter data set for experimentation. The model summarizes the collected hot topics of social media into a short sentence. The resulting summary can help users quickly understand topic trends on social media, most of which are in line with the development of real-world events. Furthermore, in order to verify the quality of the text summary results, we conducted a survey on the resulting summaries with human judges for quality assessment of generated summaries. The experimental result is discussed in detail as follows:

(1) The T5 model is trained in an end-to-end and text-to-text manner, with text as input and summary text as output. In this work, we demonstrated that T5 has achieved state-of-the-art results in many NLP benchmarks and can be fine-tuned to perform automatic summarization tasks well.

(2) As mentioned earlier, due to the real-time nature of social media, an automatic event summary system using Twitter data was developed in this work to provide people with an updated summary of the event while the situation is still evolving. However, it is worth mentioning that the "real-time" here means that the summary system can expand or update the summary at any point in time, rather than real-time computing.

(3) Going further, "real-time event summarization" is different from "real-time event detection" methods. The summarization system we developed can summarize the continuously accumulated hot topic tweets into a short sentence, instead of just performing real-time event detection. Compared to real-time event detection methods, detected events on Twitter are often short-term. The summary generated by our system can represent a simplified version of the event tweet message of interest, starting at any point in time of the ongoing event.

(4) Generally speaking, abstractive summarization models are prone to produce grammatically correct incoherent sentences. In the work of summarizing text, it is easy for human researchers to make language errors. The advantages of using abstractive summarization models enable us to alleviate some of the grammatical problems of real people. Despite the results of the system evaluation show that the summaries we got are not perfect in terms of grammatical fluency, the results are acceptable and close to the performance of human researchers in the real world.

(5) Although this developed model can quickly summarize the text, it still has certain limitations in generality. The social media data used in this study only uses English data on Twitter to generate summaries. Although English is the common language of most people in the world, it must be only through English to understand the various events in the world, there are still some prejudices. In the future, we consider using multilingual models and multilingual data sets (made by different native speakers) to train the model so that we can more accurately grasp various world events from different perspectives.

(6) In this work, we implemented a method of generating storylines using the automatic summarization technique we developed. The generated storyline is composed of sub-events or key events related to a given major event. The sub-events show the status of the main event in progress. Automated storyline generation has been a research issue since nearly the inception of artificial intelligence. In particular, the storylines of events relate to identify entities and summarize events that leads to the event of interest. However, the proposed system model has not yet performed the full function of the "automated storyline generation" paradigm for understanding event development. This is because in this work, we have not provided methods for

analyzing the evolution of specific event entities and exploring the entities and their relationships in event stories. We will leave the work of constructing event storyline for future work to complete.

7. Conclusions

The uncertainty of emergencies and natural disasters may affect companies and everyone's lives. During such high-impact events, people may want to stay up-to-date as the event develops because they want to better understand the situation and possible trend of the event. In this case, it is almost impossible for anyone to browse through all of the news articles to understand the event stories required for decision-making. Automated summarization systems that access a large number of information sources in real time can help by providing event summaries, which are updated whenever important sub-events are detected. The update of the event will help those interested persons to understand the situation of the current event.

In this work, we developed a hybrid method to combine a real-time event summarization system and a self-adaptive Twitter-based event detection method, using pre-trained language models on the tweet data set for online machine learning. Hence, each user can obtain real-time information about specific incidents (or events) and the agility to make appropriate decisions quickly. The goal of this work can be divided into the following two parts: (1). In a high-impact event, people may want to learn the latest information as the event develops. It is necessary to develop an effective real-time event summary system to obtain sufficient real-time intelligence information for effective decision-making; (2). In order to acquire extensive coverage of event details, using pre-trained language models can help us expand more knowledge and trending topics about how event stories develop. Our research is intended as a starting point for exploring the benefits of this potential for event awareness.

The novelty and contributions of the work are summarized as follows:

- We implemented an end-to-end and text-to-text summarization system model, in which text streams as input and summary text as output.
- Our system model combined a real-time event summarization system and a self-adaptive Twitter-based event detection method. The developed summarization system can summarize the continuously accumulated hot topic tweets into a short sentence, instead of just performing real-time event detection.
- The developed system used pre-trained language models to help users obtain more knowledge and trending topics about how event stories develop.
- The system can support users to obtain sufficient real-time intelligence information in specific events to make effective decisions.

For future work, it will be interesting to further incorporate the sentiment of the tweet into the summary of related events to predict the world market trends affected by the event. On the other hand, as mentioned earlier, building a summary system in the "automatic storyline generation" paradigm is very useful for understanding the details of event development. In this way, it is possible to explore the evolution of a specific event entity in the event story and the relationship between the entity and its event. In addition, by taking advantage of the potential of pre-trained language models, it will be very useful to gather intelligence on multilingual event summaries from multilingual social media messages. Finally, the use of automatic video summarization methods to extend the developed real-time text summarization methods to generate event summaries is also a promising application.

Author Contributions: Conceptualization, C.-H.L. and Y.-L.C.; methodology, C.-H.L.; software, Y.-L.C.; validation, H.-C.Y., Y.J.C. and Y.-L.C.; formal analysis, H.-C.Y.; investigation, C.-H.L.; resources, Y.J.C.; data curation, H.-C.Y.; writing—original draft preparation, C.-H.L.; writing—review and editing, C.-H.L.; visualization, Y.-L.C.; supervision, C.-H.L.; project administration, C.-H.L.; funding acquisition, C.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

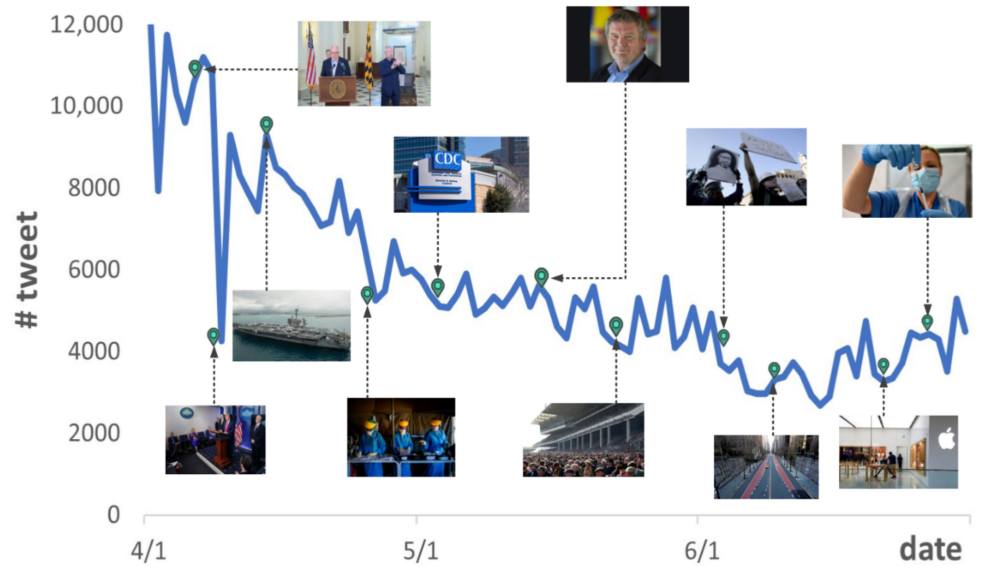
Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

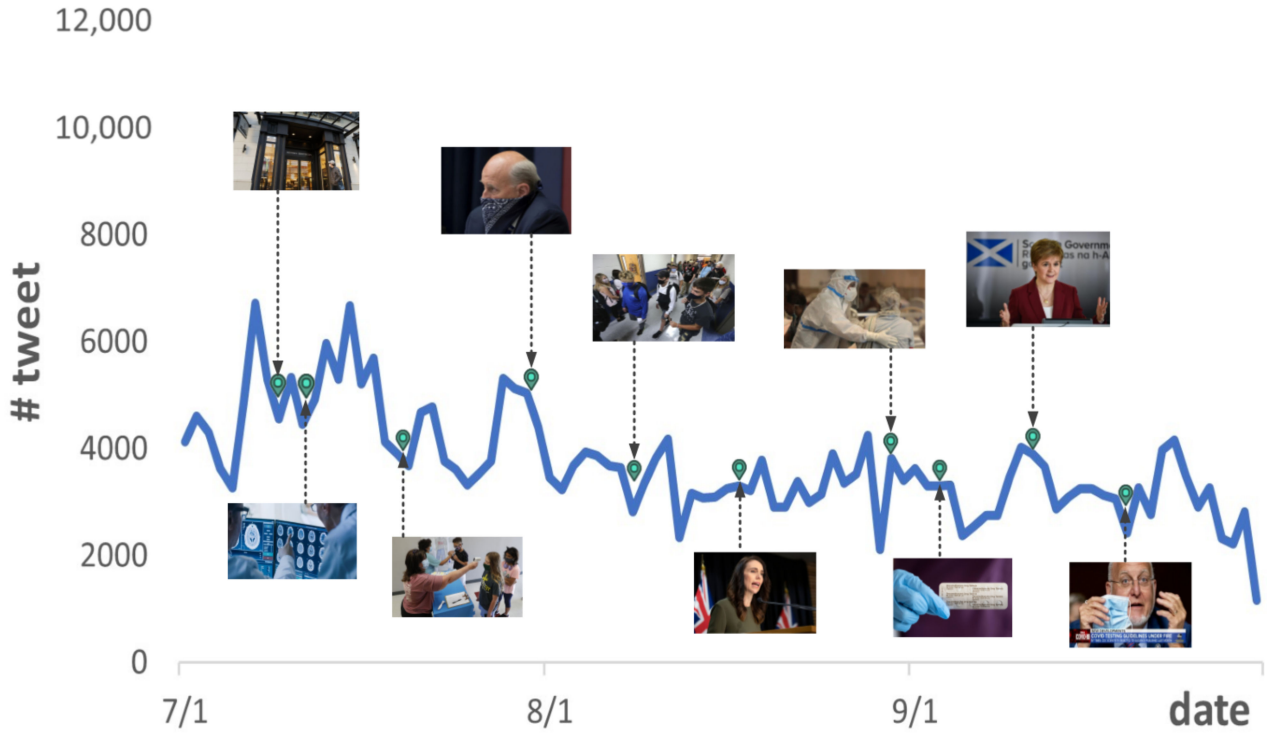
Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Examples of Generated Storyline Related to COVID-19 Pandemic



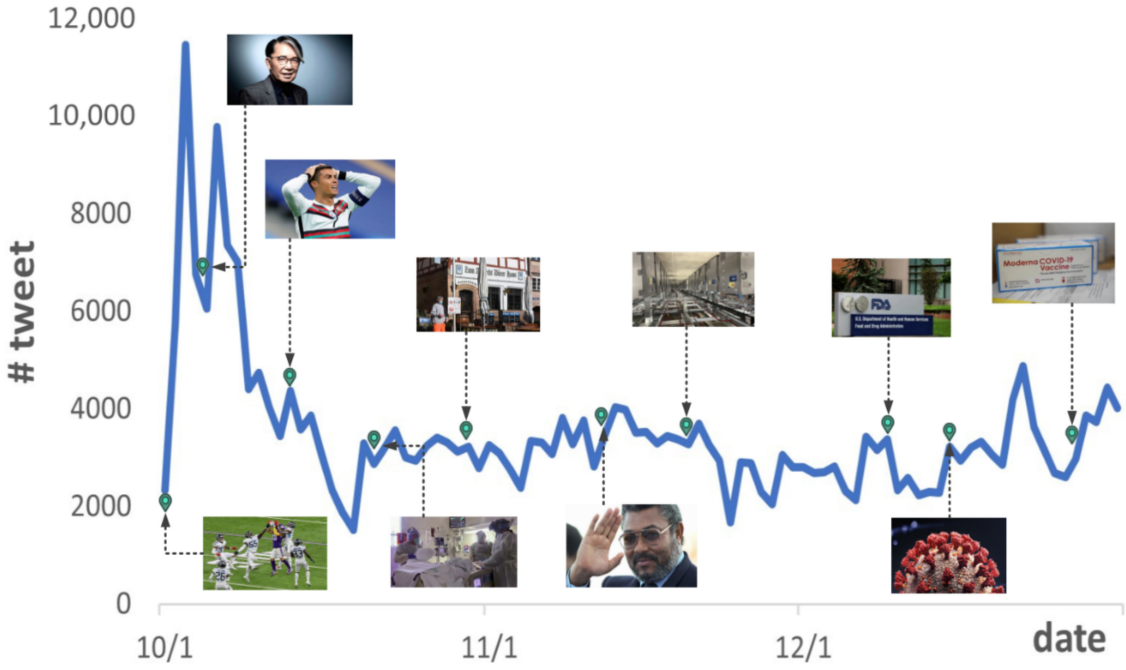
2020/4/3	Romelu Lukaku's Serie A contract has been suspended after the coronavirus pandemic was confirmed.
2020/4/4	Italy is suffering from a Coronavirus infection.
2020/4/5	A look at some of the key stories from the week leading up to the Coronavirus pandemic.
2020/4/6	The US state of Maryland has reported a rise in new cases of coronavirus, with one new death.
2020/4/7	Donald Trump is facing a major crisis in the US, with millions of people losing their jobs because of a deadly virus.
2020/4/8	The Indian government has announced a special fund to fight the deadly coronavirus in the northern state of Tamil Nadu.
2020/4/9	US spy agencies are tracking the rise of a novel coronavirus in China as early as November, according to reports.
2020/4/10	Live coverage of the Coronavirus pandemic in the US.
2020/4/11	Apple and Google are collaborating on technology to help track the spread of the Coronavirus (COVID).
2020/4/12	China?â?propaganda doctor has released a video showing the death of a woman from a virus made in China.
2020/4/13	The first US sailor to die from Coronavirus after an outbreak on the USS Theodore Roosevelt has died.
2020/4/14	The World Health Organization (WHO) has released a list of the world's most deadly Coronaviruses.
2020/4/15	The Coronavirus Pandemic is causing the deaths of more than 11,000 people in the US and Canada.
2020/4/16	The death of a Nigerian nurse after contracting the Coronavirus is the latest in a series of news stories from around the world.
2020/4/17	The US government is stepping up testing of a new vaccine against Coronavirus NCDC, according to preliminary data.
⋮	
2020/6/4	The death of George Floyd in Gombe state, West Bengal, was caused by a coronavirus.
2020/6/8	Oil giant BP has announced plans to cut global workforce by jobs amid a pandemic of COVID.
2020/6/9	Lockdowns in the US helped prevent millions of Coronavirus Infections, a study suggests.
2020/6/11	The death toll from the Coronavirus pandemic could have been reduced by half if the government had introduced lockdown earlier, says a former senior scientist.
2020/6/12	A look at some of the most interesting facts about China's pandemic.
2020/6/16	NFL players and coaches have tested positive for coronavirus.
2020/6/17	The Chinese virus has been linked to China, according to a report by the BBC.
2020/6/18	The Republic of Ireland's Coronavirus outbreak has been declared a national emergency.
2020/6/20	Apple is closing some stores in the US as cases of Covid virus (COVID) rise.
2020/6/22	China halts imports of poultry from the United States over concerns about an outbreak of Covid virus.
2020/6/24	The tennis star, Novak Djokovic, has tested positive for coronavirus.
2020/6/25	A new vaccine for Coronavirus has been trialled in the UK.
2020/6/28	The US state of Florida has been closed to new cases of coronavirus, the vice president's office says.
2020/6/29	The death toll from the Coronavirus (COVID) pandemic has risen to five million, according to new figures.
2020/6/30	China is a virus factory, but it's also one of the world's biggest.

Figure A1. Generated storyline related to COVID-19 from April to June 2020.



2020/7/1	China is a country of demons.
2020/7/3	A global record for the number of confirmed cases of Coronavirus (COVID) in the US has been broken as more cases were confirmed across the country on Thursday.
2020/7/6	A new virus has been discovered in China.
2020/7/7	China has confirmed a case of bubonic plague in its Inner Mongolia region.
2020/7/8	US President-elect Donald Trump has vowed to shut down schools in the wake of the Coronavirus pandemic.
2020/7/9	US retailer Brooks Brothers has filed for bankruptcy protection, citing pandemic struggles and plans to close stores.
2020/7/10	China warns of a new pneumonia that is potentially deadlier than the novel coronavirus Covid in Kazakhstan.
2020/7/12	A new report from the US National Institutes of Health says there is a growing divide in the field of coronavirus.
2020/7/13	What's the difference between a corona virus and a Chinese product?
2020/7/14	BBC News NI examines the latest developments in the fight against the coronavirus.
2020/7/15	Pls spread the truth about ccp virus in Hong Kong.
2020/7/16	The US government is not doing enough to stop the spread of pandemics, but a new reporting system that requires hospitals to send patient data to the Centers for Disease Control (CDC) has been put on hold.
2020/7/19	News from around the world on Wednesday as infants tested positive for Coronavirus (COVID) in one Texas county.
2020/7/20	The UK is to suspend its extradition treaty with Hong Kong amid rising tensions with China.
2020/7/21	Oregon Governor Tom Ford has said he is "shocked" by the state's failure to restore school safety.
⋮	⋮
2020/8/18	A former CIA officer has been arrested on charges that he conspired with a relative who served in the US military to spy for China.
2020/8/30	India's coronavirus (COVID) tally has reached its highest level since the pandemic began in January with two provinces.
2020/8/31	The US Centers for Disease Control and Prevention (CDC) has updated its Covid Hoax death rate to admit that only one of the deaths was caused by the virus.
2020/9/1	The death toll from the Coronavirus (COVID) pandemic in the US has risen to more than million, according to health officials.
2020/9/3	Steroids are being used to treat the Coronavirus (COVID) and other serious diseases.
2020/9/4	The Batman filming has been suspended in the UK after a positive test for COVID.
2020/9/6	Bill Gates-funded polio Vaccine is Causing A Deadly Outbreak In Africa, the UN has said.
2020/9/11	BBC News NI is to stop broadcasting all of Nicola Sturgeon's coronavirus briefings.
2020/9/16	Kim Kardashian West and other celebrities are to freeze their social media accounts in protest at Facebook's inaction over hate speech.
2020/9/19	The US Centers for Disease Control and Prevention (CDC) has reversed a recommendation that said people without symptoms should not be tested for the coronavirus.
2020/9/20	The Chinese Communist Party (CCP) is threatening to destroy the world, and it has spread the corona virus.
2020/9/21	The US Centers for Disease Control and Prevention (CDC) has updated its guidance on the spread of the Coronavirus.
2020/9/22	The US Centers for Disease Control (CDC) says it is updating its guidelines on airborne transmission of the coronavirus.
2020/9/23	China virus.
2020/9/24	A massive genetic study has shown that the coronavirus is mutating and potentially evolving amid rapid US spread via.

Figure A2. Generated storyline related to COVID-19 from July to September 2020.



2020/10/1	The National Football League (NFL) has postponed two games this weekend due to a COVID pandemic.
2020/10/2	The Chinese virus.
2020/10/4	The NFL's two biggest games of the season have been postponed after Cam Newton tested positive for coronavirus.
2020/10/5	Kenzo Takada, the founder of the Japanese fashion brand Kenzo Takada, has died aged 81 from COVID disease.
2020/10/8	China's communist party virus has been blamed for a number of attacks on the US.
2020/10/9	China virus.
2020/10/10	A look at the latest developments in the Chinese Virus.
2020/10/13	Cristiano Ronaldo has tested positive for the Covid virus, Portugal's football federation says.
2020/10/14	Dustin Johnson withdraws from the CJ Cup after a positive coronavirus test.
2020/10/15	Alabama head coach Nick Saban has tested positive for Coronavirus (COVID) after being told he would not be allowed to play football.
2020/10/17	It's made in China.
2020/10/18	In this series of letters from African journalists, novelist and writer Li-Meng Yan explains why the coronavirus is so bad in China.
2020/10/21	The US Centers for Disease Control and Prevention (CDC) says it has caused more deaths than expected during the Covid pandemic.
⋮	⋮
2020/12/9	The US Food and Drug Administration (FDA) has approved a new vaccine for the COVID virus.
2020/12/10	The European Medicines Agency has launched an investigation into a cyberattack that has seen data about a Covid vaccine stolen.
2020/12/13	Nigeria's governor Babajide Sanwo-Olu has tested positive for Coronavirus (COVID) in the state of Lagos.
2020/12/15	A new strain of coronavirus has emerged in the UK, according to a report by BBC News.
2020/12/16	In the wake of a report by Farrakhan on the dangers of vaccines, we take a look at some of the things you should know about the vaccine
2020/12/17	US Secretary of State Mike Pompeo has said a health worker in Alaska had an allergic reaction to the COVID vaccine.
2020/12/18	The US government has halted deliveries of a vaccine for the Covid virus, saying it will not be available until next year.
2020/12/20	Marco Rubio has been given the Covid vaccine, a new strain of the Coronavirus.
2020/12/21	The BBC's Jeremy Corbyn has been talking about the latest outbreak of the Coronavirus in the US.
2020/12/27	A US doctor has reported a severe allergic reaction to the Moderna COVID vaccine.
2020/12/28	The US Centers for Disease Control and Prevention (CDC) has issued new guidance on the COVID vaccine.
2020/12/29	Everton's match against Manchester City has been postponed because of a growing outbreak of Covid.
2020/12/30	The Covid vaccine has been approved for use in the UK after a new strain of coronavirus was discovered at an Oxford University hospita

Figure A3. Generated storyline related to COVID-19 from October to December 2020.

Appendix B. Example Model Outputs: Tweet Summaries of the "COVID-19 Pandemic" Sub-Events

Figure A4 shows that on 24 January 2020, the World Health Organization issued a global warning about the new coronavirus that was discussed in social media.

2020/1/24 #80

Document :

- How to protect yourself from deadly new coronavirus CNN Video
- IHR Emergency Committee continues to monitor novel coronavirus n-CoV
- Mapping the spread of the new coronavirus
- there's a new strain of the corona virus its called the modelovirus.
- The coronavirus I want to get
- the coronavirus supposedly
- The new coronavirus Symptoms prevention and what to do if you think you're infected
- How the new coronavirus compares to Ebola Zika and SARS Washington Post
- Preparing for the Novel Coronavirus
- Wuhan novel coronavirus what you need to know
- National WHO issues worldwide warning on new coronavirus
- Hopefully has coronavirus
- Anybody paying attention to that new carona virus that's being spread
- Is this where that new CoronaVirus is coming from jeez
- ways to protect yourself from coronavirus according to an infectious diseases expe ABS-CBN News

SUMMARY :

The World Health Organization (WHO) has issued a worldwide warning on the new coronavirus.

Figure A4. A sample resulting summary of formulated event topics (24 January 2020).

Figure A5 shows that on 12 February 2020, the World Health Organization officially named the new coronavirus COVID and was discussed in social media.

2020/2/12 #10

Document :

- Coronavirus officially named Covid says WHO
- Here's one of the SARS papers on transmission Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus NEJM
- Coronavirus gets a name COVID
- Covid The WHO name for the Novel corona virus
- The World Health Organisation has confirmed UN agencies have decided to assign a new name to the novel coronavirus
- New name for coronavirus is covid
- WHO officially names the illness caused by new coronavirus COVID
- Preparation for possible sustained transmission of novel coronavirus Lessons from previous epide
- At least Wuhan medical staff infected with coronavirus
- CoronaVirus has been given a formal name by the World Health Organization (COVID
- World Health Organization officially names Coronavirus Covid
- The name provides a standard format for future coronavirus outbreaks the WHO says
- Disease Caused by the Novel Officially Has a Name COVID via
- According to the organization the word refers to a group of viruses that contains those that cause SARS severe acute respiratory syndrome and MERS Middle East respiratory syndrome
- The World Health Organization says the official name for the disease caused by the new coronavirus is Covid Foreign Media
- WHO Names Coronavirus Disease As COVID
- WHO has named Coronavirus as COVID
- Coronaviruses Symptoms treatments and variants

⋮

SUMMARY :

The World Health Organization (WHO) has officially named the novel coronavirus Covid.

Figure A5. A sample resulting summary of formulated event topics (12 February 2020).

Figure A6 illustrates the topics that will be postponed for the 2020 Olympic Games discussed on 30 March 2020.

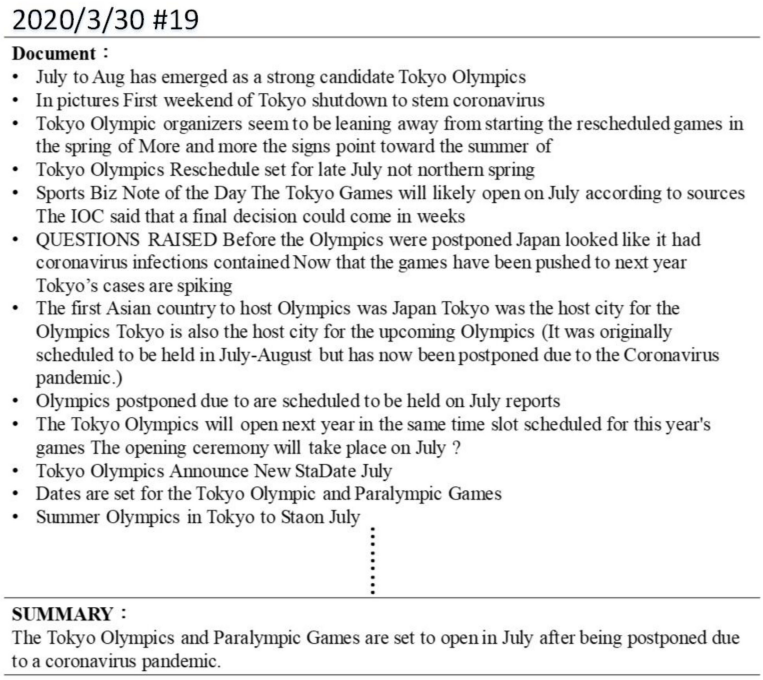


Figure A6. A sample resulting summary of formulated event topics (30 March 2020).

Figure A7 shows that the possible variants of the new coronavirus were discussed in social media on 6 May 2020, and they may become more toxic and more contagious than the original.

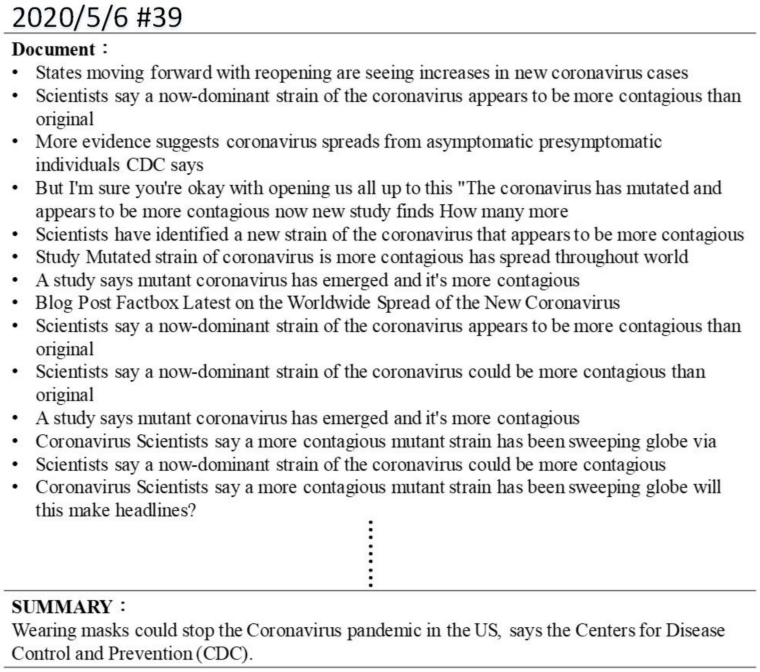
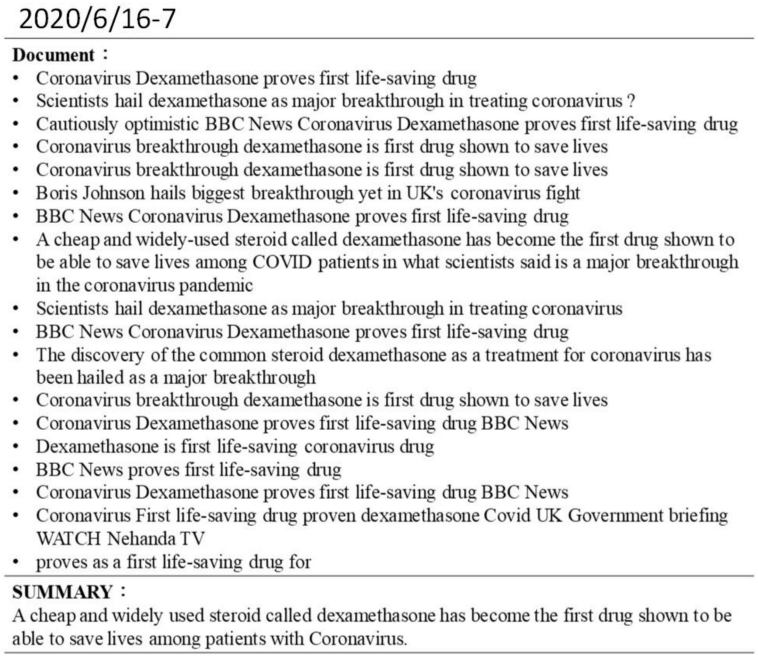


Figure A7. A sample resulting summary of formulated event topics (6 May 2020).

Figure A8 shows that the Dexamethasone drug was discussed in social media on 17 June 2020 to help save the lives of patients with the new coronavirus.


2020/6/16-7

Document :

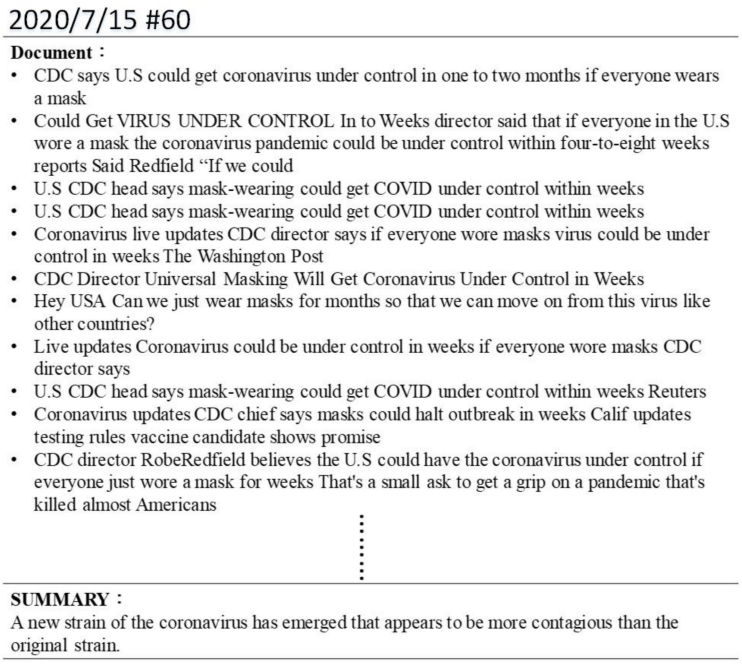
- Coronavirus Dexamethasone proves first life-saving drug
- Scientists hail dexamethasone as major breakthrough in treating coronavirus ?
- Cautiously optimistic BBC News Coronavirus Dexamethasone proves first life-saving drug
- Coronavirus breakthrough dexamethasone is first drug shown to save lives
- Coronavirus breakthrough dexamethasone is first drug shown to save lives
- Boris Johnson hails biggest breakthrough yet in UK's coronavirus fight
- BBC News Coronavirus Dexamethasone proves first life-saving drug
- A cheap and widely-used steroid called dexamethasone has become the first drug shown to be able to save lives among COVID patients in what scientists said is a major breakthrough in the coronavirus pandemic
- Scientists hail dexamethasone as major breakthrough in treating coronavirus
- BBC News Coronavirus Dexamethasone proves first life-saving drug
- The discovery of the common steroid dexamethasone as a treatment for coronavirus has been hailed as a major breakthrough
- Coronavirus breakthrough dexamethasone is first drug shown to save lives
- Coronavirus Dexamethasone proves first life-saving drug BBC News
- Dexamethasone is first life-saving coronavirus drug
- BBC News proves first life-saving drug
- Coronavirus Dexamethasone proves first life-saving drug BBC News
- Coronavirus First life-saving drug proven dexamethasone Covid UK Government briefing WATCH Nehanda TV
- proves as a first life-saving drug for

SUMMARY :

A cheap and widely used steroid called dexamethasone has become the first drug shown to be able to save lives among patients with Coronavirus.

Figure A8. A sample resulting summary of formulated event topics (17 June 2020).

Figure A9 shows that on 15 July 2020, in social media, the Center for Disease and Control of the United States called on people to wear masks to effectively prevent the new coronavirus.


2020/7/15 #60

Document :

- CDC says U.S could get coronavirus under control in one to two months if everyone wears a mask
- Could Get VIRUS UNDER CONTROL In to Weeks director said that if everyone in the U.S wore a mask the coronavirus pandemic could be under control within four-to-eight weeks reports Said Redfield "If we could
- U.S CDC head says mask-wearing could get COVID under control within weeks
- U.S CDC head says mask-wearing could get COVID under control within weeks
- Coronavirus live updates CDC director says if everyone wore masks virus could be under control in weeks The Washington Post
- CDC Director Universal Masking Will Get Coronavirus Under Control in Weeks
- Hey USA Can we just wear masks for months so that we can move on from this virus like other countries?
- Live updates Coronavirus could be under control in weeks if everyone wore masks CDC director says
- U.S CDC head says mask-wearing could get COVID under control within weeks Reuters
- Coronavirus updates CDC chief says masks could halt outbreak in weeks Calif updates testing rules vaccine candidate shows promise
- CDC director RobeRedfield believes the U.S could have the coronavirus under control if everyone just wore a mask for weeks That's a small ask to get a grip on a pandemic that's killed almost Americans

SUMMARY :

A new strain of the coronavirus has emerged that appears to be more contagious than the original strain.

Figure A9. A sample resulting summary of formulated event topics (15 July 2020).

Figure A10 shows that it was discussed in social media on 16 November 2020. American drug maker Moderna stated that its COVID vaccine can effectively prevent the coronavirus.

2020/11/16 #0

Document :

- Ugur Sahin who helped develop Pfizer's Covid vaccine is confident it will work
- The Truth About the Coronavirus Vaccine Trials
- *MODERNA'S COVID VACCINE CANDIDATE MEETS PRIMARY EFFICACY
- *MODERNA CORONAVIRUS VACCINE FOUND EFFECTIVE IN ANALYSIS
- BREAKING vaccine is effective
- Breaking Early data shows Moderna's coronavirus vaccine is effective in preventing disease The second mRNA jab to return such promising results after Pfizer's Full story below
- BREAKING Moderna coronavirus vaccine found to be effectiv
- A Covid vaccine developed in the USA has been found to be effective:
- told you so on the right track COVID Moderna vaccine shown to be effective and easier to store according to interim analysis
- Moderna Says Its Covid Vaccine Was Effective in Latest Trial The Wall Street Journal Moderna Says Its Covid Vaccine Was Effective in Latest Trial
- Moderna's coronavirus vaccine is effective according to company data
- Moderna's coronavirus vaccine is effective according to company data
- Moderna's Covid Vaccine Is Effective in Early Results Firm Says The Wall Street Journal Outstanding
- BREAKING Moderna coronavirus vaccine found to be effective
- Moderna Says Its Vaccine Is Effective in Preventing COVID via
- Second Coronavirus Vaccine Nearly Percent Effective in Preliminary Study via
- Moderna announces Covid vaccine breakthrough

SUMMARY :

The US drugmaker Moderna says its Covid vaccine is effective in preventing the Coronavirus.

Figure A10. A sample resulting summary of formulated event topics (16 November 2020).

Figure A11 shows that Pfizer's COVID vaccine was discussed in social media on 9 December 2020, in compliance with the safety and effectiveness guidelines for emergency use.

2020/12/9 #12

Document :

- The Washington Post FDA review confirms safety and efficacy of Pfizer coronavirus vaccine
- The FDA's vaccine advisory committee released a detailed analysis finding that Pfizer's coronavirus vaccine appears to meet the safety and efficacy requirements necessary for an emergency use authorization
- US regulators post positive review of Pfizer vaccine data
- Rachel Feltman The FDA says Pfizer's COVID vaccine is effective and safe
- What does effective mean for Pfizer's Covid vaccine
- US regulators post positive review of Pfizer vaccine data via
- The result of the COVID Vaccine
- The FDA says Pfizer's COVID vaccine is effective and safe
- US regulators post positive review of Pfizer vaccine data
- Good explained on how the COVID vaccine works
- Pfizer CEO confident of getting U.S advisory panel nod for COVID vaccine
- Regulatory approval of Pfizer BioNTech vaccine for COVID
- FDA Pfizer Covid vaccine data fits with guidance on emergency authorization
- FDA Analysis Confirms That Pfizer/BioNTech COVID Vaccine May Be Add your highlights
- Regulatory approval of Pfizer BioNTech vaccine for COVID
- "Pfizer's COVID vaccine is safe and effective according to the FDA
- U.S regulators post positive review of Pfizer vaccine data
- FDA review confirms safety and efficacy of Pfizer coronavirus vaccine
- Health Canada approves Pfizer COVID vaccine

SUMMARY :

The US Food and Drug Administration (FDA) has confirmed that Pfizer's COVID vaccine meets safety and efficacy guidelines for emergency use.

Figure A11. A sample resulting summary of formulated event topics (9 December 2020).

References

1. Rudrapal, D.; Das, A.; Bhattacharya, B. A survey on automatic Twitter event summarization. *J. Inf. Process. Syst.* **2018**, *14*, 79–100.
2. Chakrabarti, D.; Punera, K. Event summarization using tweets. In Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5.
3. Chua, F.C.T.; Asur, S. Automatic summarization of events from social media. In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, USA, 8–11 July 2013.
4. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv* **2019**, arXiv:1910.10683.
5. Lee, C.-H. Mining spatio-temporal information on microblogging streams using a density-based online clustering method. *Expert Syst. Appl.* **2012**, *39*, 9623–9641. [CrossRef]
6. Lee, C.-H.; Chien, T.-F. Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking. *J. Inf. Sci.* **2013**, *39*, 523–543. [CrossRef]
7. Sharifi, B.; Hutton, M.A.; Kalita, J. Summarizing microblogs automatically. In Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, CA, USA, 2–4 June 2010; pp. 685–688.
8. Inouye, D.; Kalita, J.K. Comparing Twitter Summarization Algorithms for Multiple Post Summaries. In Proceedings of the 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011; pp. 298–306.
9. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
10. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
11. Rachabathuni, P.K. A survey on abstractive summarization techniques. In Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 23–24 November 2017; pp. 762–765.
12. Gambhir, M.; Gupta, V. Recent automatic text summarization techniques: A survey. *Artif. Intell. Rev.* **2017**, *47*, 1–66. [CrossRef]
13. Neto, J.L.; Freitas, A.A.; Kaestner, C.A.A. Automatic Text Summarization Using a Machine Learning Approach. In Proceedings of the 16th Brazilian Symposium on Artificial Intelligence, Porto de Galinhas/Recife, Brazil, 11–14 November 2002; pp. 205–215.
14. Tas, O.; Kiyani, F. A survey automatic text summarization. *J. PressAcademia Procedia* **2017**, *5*, 205–213. [CrossRef]
15. Radev, D.R.; Allison, T.; Blair-Goldensohn, S.; Blitzer, J.; Celebi, A.; Dimitrov, S.; Drabeck, E.; Hakim, A.; Lam, W.; Liu, D.; et al. MEAD-a platform for multidocument multilingual text summarization. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon, Portugal, 26–28 May 2004; pp. 699–702.
16. Nenkova, A.; Vanderwende, L.; McKeown, K. A compositional context sensitive multi-document summarizer: Exploring the factors that influence summarization. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 6–11 August 2006; pp. 573–580.
17. Filatova, E.; Hatzivassiloglou, V. Event-based extractive summarization. In Proceedings of the ACL-04 Workshop, Association for Computational Linguistics, Barcelona, Spain, 21–26 July 2004; pp. 104–111.
18. Moratanch, N.; Chitrakala, S. A survey on extractive text summarization. In Proceedings of the 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, 10–11 January 2017; pp. 1–6.
19. Mutlu, B.; Sezer, E.A.; Akcayol, M.A. Multi-document extractive text summarization: A comparative assessment on features. *J. Knowl.-Based Syst.* **2019**, *183*, 104848. [CrossRef]
20. Kupiec, J.; Pedersen, J.; Chen, F. A trainable document summarizer. In Proceedings of the 18th Annual International ACM SIGIR conference on Research and Development in Information Retrieval–SIGIR '95, Seattle, WA, USA, 9–13 July 1995; pp. 68–73.
21. Conroy, J.M.; O'Leary, D.P. Text summarization via hidden Markov models. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR '01, New Orleans, LA, USA, 9–13 September 2001; pp. 406–407.
22. Woodsend, K.; Lapata, M. Automatic generation of story highlights. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11–16 July 2010; pp. 565–574.
23. Isonuma, M.; Fujino, T.; Mori, J.; Matsuo, Y.; Sakata, I. Extractive Summarization Using Multi-Task Learning with Document Classification. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 2101–2110.
24. Nallapati, R.; Zhai, F.; Zhou, B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
25. Gunawan, D.; Pasaribu, A.; Rahmat, R.F.; Budiarto, R. Automatic Text Summarization for Indonesian Language Using TextTeaser. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing Ltd.: Semarang, Indonesia, 2016; Volume 190, p. 012048.
26. Narayan, S.; Cohen, S.B.; Lapata, M. Ranking Sentences for Extractive Summarization with Reinforcement Learning. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, LA, USA, 1–6 June 2018; pp. 1747–1759.

27. Dong, Y.; Shen, Y.; Crawford, E.; Van Hoof, H.; Cheung, J.C.K. BanditSum: Extractive Summarization as a Contextual Bandit. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3739–3748.
28. Rush, A.M.; Chopra, S.; Weston, J. A Neural Attention Model for Abstractive Sentence Summarization. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 379–389.
29. Chopra, S.; Auli, M.; Rush, A.M. Abstractive Sentence Summarization with Attentive Recurrent Neural Networks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 93–98.
30. Song, S.; Huang, H.; Ruan, T. Abstractive text summarization using LSTM-CNN based deep learning. *J. Multimed. Tools Appl.* **2018**, *78*, 857–875. [CrossRef]
31. Paulus, R.; Xiong, C.; Socher, R. A deep reinforced model for abstractive summarization. *arXiv* **2017**, arXiv:1705.04304.
32. Chen, Y.-C.; Bansal, M. Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 675–686.
33. Masum, A.K.M.; Abujar, S.; Talukder, A.I.; Rabby, A.S.A.; Hossain, S.A. Abstractive method of text summarization with sequence to sequence RNNs. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; pp. 1–5.
34. Cai, T.; Shen, M.; Peng, H.; Jiang, L.; Dai, Q. Improving Transformer with Sequential Context Representations for Abstractive Text Summarization. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Dunhuang, China, 9–14 October 2019; pp. 512–524.
35. Qiu, X.; Sun, T.; Xu, Y.; Shao, Y.; Dai, N.; Huang, X. Pre-trained models for natural language processing: A survey. *J. Sci. China Technol. Sci.* **2020**, *63*, 1872–1897. [CrossRef]
36. Liu, Y.; Lapata, M. Text Summarization with Pretrained Encoders. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3721–3731.
37. Kieuvongngam, V.; Tan, B.; Niu, Y. Automatic text summarization of Covid-19 medical research articles using bert and gpt-2. *arXiv* **2020**, arXiv:2006.01997.
38. Khandelwal, U.; Clark, K.; Jurafsky, D.; Kaiser, L. Sample efficient text summarization using a single pre-trained transformer. *arXiv* **2019**, arXiv:1905.08836.
39. Zhang, X.; Wei, F.; Zhou, M. HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5059–5069.
40. Farahani, M.; Gharachorloo, M.; Manthouri, M. Leveraging ParsBERT and Pretrained mT5 for Persian Abstractive Text Summarization. In Proceedings of the 2021 26th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 3–4 March 2021; pp. 1–6.
41. Ma, T.; Pan, Q.; Rong, H.; Qian, Y.; Tian, Y.; Al-Nabhan, N. T-BERTSum: Topic-Aware Text Summarization Based on BERT. *J. IEEE Trans. Comput. Social Syst.* **2021**, 1–12. [CrossRef]
42. Kerui, Z.; Haichao, H.; Yuxia, L.; Zhang Kerui Lazada Group, China; Hu Haichao SF Technology, Shenzhen China; Liu Yuxia SF Technology, Shenzhen China. Automatic Text Summarization on Social Media. In Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control, Newcastle upon Tyne, UK, 17–19 November 2020; pp. 1–5.
43. Garg, A.; Adusumilli, S.; Yenneti, S.; Badal, T.; Garg, D.; Pandey, V.; Nigam, A.; Gupta, Y.K.; Mittal, G.; Agarwal, R. NEWS Article Summarization with Pretrained Transformer. In *Communications in Computer and Information Science, Proceedings of the Advanced Computing, IACC 2020, Panaji, Goa, India, 5–6 December 2020*; Springer: Singapore, 2020; pp. 203–211. [CrossRef]
44. Daiya, D. Combining Temporal Event Relations and Pre-Trained Language Models for Text Summarization. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 14–17 December 2020; pp. 641–646.
45. Fecht, P.; Blank, S.; Zorn, H.P. Sequential Transfer Learning in NLP for German Text Summarization. In Proceedings of the 4th edition of the Swiss Text Analytics Conference, Winterthur, Switzerland, 18–19 June 2019.
46. Narayan, S.; Cohen, S.B.; Lapata, M. Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 1797–1807.
47. Lee, C.-H.; Wu, C.-H.; Chien, T.-F. BursT: A Dynamic Term Weighting Scheme for Mining Microblogging Messages. *J. Int. Symp. Neural Netw.* **2011**, *6677*, 548–557. [CrossRef]
48. Lee, C.-H.; Wu, C.-H. A Self-adaptive Clustering Scheme with a Time-Decay Function for Microblogging Text Mining. *J. Inf. Technol.* **2011**, *185*, 62–71. [CrossRef]
49. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In Proceedings of the 27th Interna-tional Conference on Neural Information Processing Systems, Montréal, QC, Canada, 6–12 December 2014; pp. 3104–3112.

50. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, LA, USA, 1–6 June 2018; Volume 1, pp. 2227–2237.
51. Howard, J.; Ruder, S. Universal Language Model Fine-tuning for Text Classification. *arXiv* **2018**, arXiv:1801.06146.
52. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.10084.
53. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain, 21–26 July 2004; pp. 74–81.
54. Shigemura, J.; Ursano, R.J.; Morganstein, J.C.; Kurosawa, M.; Benedek, D.M. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: Mental health consequences and target populations. *J. Psychiatry Clin. Neurosci.* **2020**, *74*, 281–282. [CrossRef] [PubMed]