CS 4984 / CS 5984
Big Data Text Summarization
Taught by Edward A. Fox

Virginia Tech
Blacksburg, VA 24061

Abstractive Text Summarization of the Parkland Shooting Collection

Ryan Kingery

Jiacheng Ye

Li Jun Huang

Chao Xu

Sudha Ravali Yellapantula

Submitted on December 12, 2018
Funded by NSF: IIS-1619028

Contents

List of Figures

List of Tables

Abstract

We analyze various ways to perform abstractive text summarization on an entire collection of news articles. We specifically seek to summarize the collection of web-archived news articles relating to the 2018 shooting at Marjory Stoneman Douglas High School in Parkland, Florida. The original collection contains about 10,100 archived web pages that mostly relate to the shooting, which after pre-processing reduces to about 3,900 articles that directly relate to the shooting.

We then explore several ways to generate abstractive summaries for the collection using deep learning methods. Since current deep learning methods for abstract summarization are only capable of summarizing text at the single-article level or below, to perform summarization on our collection, we identify a set of representative articles from the collection, summarize each of those articles using our deep learning models, and then concatenate those summaries together to produce a summary for the entire collection.

To identify the representative articles to summarize we investigate various unsupervised methods to partition the space of articles into meaningful groups. We try choosing these articles by random sampling from the collection, by using topic modeling, and by sampling from clusters obtained from clustering on Doc2Vec embeddings. To summarize each individual article we explore various state of the art deep learning methods for abstractive summarization: a sequence-to-sequence model, a pointer generator network, and a reinforced extractor-abstractor network.

To evaluate the quality of our summaries we employ two methods. The first is a subjective method, where each person subjectively ranked the quality of each summary. The second is an objective method which used various ROUGE metrics to compare each summary to an independently-generated gold standard summary. We found that most ROUGE scores were pretty low overall, with only the pointer-generator network on random articles picking up a ROUGE score above 0.15. This suggests that such deep learning techniques still have a lot of room for improvement if they are to be viable for collection summarization.

Chapter 1

Introduction

1.1 Introduction

Summarizing a collection of documents is a task still mostly performed by humans, and is often a task requiring large amounts of time and skill. Being able to automate such a task would likely save many organizations a great deal in resources, and would enable efficient summaries of collections of text that would be far too difficult for humans to summarize effectively. One such use case is the summarization of archived web collections. Such collections are extremely difficult to wade through, and many organizations (e.g., news organizations and universities) have a vested interest in being able to efficiently determine what important information is contained in a given archived web collection.

Ways to automate the task of text summarization have been sought after for years and have had only limited success. Part of the problem is the inherent difficulty. It is believed that text summarization is an AI-complete problem, meaning that building AI to perform text summarization at the level of humans would require that computers be as intelligent as humans. Rather than focus on this more difficult task of human-level text summarization, called abstractive text summarization, the vast majority of research over the years has instead focused on extractive text summarization, which is the simpler task of extracting the key words and phrases from a piece of text.

Only in recent years with the resurgence of deep learning techniques has

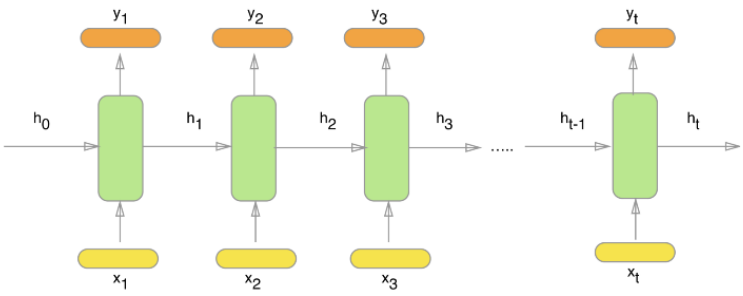

Figure 1.1: Illustration of a simple RNN. Inputs x_t map sequentially through a hidden state h_t to produce outputs y_t.

abstractive summarization started to become even remotely doable. Deep learning uses neural networks, which are cascades of multiple layers of non-linear functions, to learn representations in data. Typically, the neural networks models used to deal with text are variants of the recurrent neural network (RNN). RNNs are ways to allow neural networks to capture temporal dependencies in data and use those dependencies to learn data representations. An illustration of a simple RNN is shown in Figure 1.1. A time series $x_1, \cdots, x_t$ of input data is passed into a series of non-linear hidden units $h_1, \cdots, h_t$ to produce outputs $y_1, \cdots, y_t$.

Unlike simple neural networks, the hidden units of an RNN are allowed to depend on previous input vectors as well, i.e. $h_t = f(x_1, \cdots, x_{t-1})$, which allows them to capture temporal relationships in the data. In practice, problems arise when using simple RNNs due to their difficulty in capturing long-term temporal dependencies in data, so various extensions are used instead. The most common RNN variants used are the long-term short-term memory unit (LSTM) and the gated recurrent unit (GRU). Both work by making subtle modifications to the hidden units to make them better able to capture long-term temporal dependencies.

Using deep learning to generate text usually involves the use of sequence-to-sequence (Seq2Seq) models, which are models that take input sequences (e.g., text) and produce output sequences (e.g., more text). Seq2Seq models are built by stacking RNNs into two segments, an encoder layer and a decoder layer. The encoder layer takes the input sequence and converts it

into an internal representation. This is then passed into the decoder layer, which converts the internal representation into a sequence of outputs. Such a model can be used for text summarization, where the input sequence is a document from the text corpus, and the output sequence is the generated summary for that document. In this report, we use various extensions to Seq2Seq models to perform abstractive summarization. These extensions are described in more detail below.

1.2 Related Work

To the best of our knowledge, little if any research has been performed on generating abstract summaries for whole corpora of text. Since the resurgence of deep learning, much of the research has been focused on sentence, paragraph, and document summarization. This likely has to do with the difficulties involved in capturing long-term dependencies in larger collections of text, and with hardware limitations that arise when dealing with sequences that are too large. One of the first approaches to document summarization was based on a simple Seq2Seq model in [2]. Further extensions were made by allowing the use of extraction to deal with out-of-vocabulary words. The most successful example is the pointer-generator network of [4]. Further extensions have been made recently in [3], [5], and other sources.

Chapter 2

Models

2.1 Doc2Vec Clustering

In this section we describe Doc2Vec, an unsupervised technique that learns continuous distributed vector representations of documents [1], along with how clustering can be performed on a set of Doc2Vec embeddings to semantically segment documents in a corpus of text. Doc2Vec is based off of recent work in learning vector representations of words. A popular example of such a word vector representation that is closely related to Doc2Vec is Word2Vec [6].

The objective of a word vector representation is to build a model that uses the context of a word, i.e., the words nearby in the sentence, to predict the word itself. That is, given a word w_t occurring at position t in a sentence, the goal is to build a model to infer the probability a word w_t occurs given its context $w_{t-k}, \cdots, w_{t-1}, w_{t+1}, \cdots, w_k$. Prediction is then performed by minimizing the negative log-likelihood loss

$$loss = -\frac{1}{T}\sum_{k}^{T-k} p(w_t|w_{t-k}, \cdots, w_k).$$

Prediction is performed via the softmax function,

$$p(w_t|w_{t-k}, \cdots, w_k) = \text{softmax}(w_{t-k}, \cdots, w_k) = \frac{e^{y_{w_t}}}{\sum_i e^{y_i}},$$

where y_i is the unnormalized log-probability for each output word w_i. In the case of Word2Vec each y_i is obtained by training a single-hidden-layer neural

network to estimate the parameters U, W, b of the equation

$$y_i = b + Uh(w_{t-k}, \cdots, w_k|W),$$

where h is constructed by a concatenation or average of word vectors extracted from W. Once the model is trained, one can use the i^{th} column of W to represent the word w_i. This column vector is referred to as the embedding vector for w_i.

It has been shown in [6] and other sources that such a model is capable of capturing the semantic relationship between words. Words that are similar in meaning tend to be grouped together in the space of embedding vectors, while words with differing meanings tend to be farther away from each other. For example, words like "strong" and "powerful" are similar to each other (via cosine similarity), while "strong" and "Paris" are farther apart.

Doc2Vec extends the approach of Word2Vec to documents by essentially allowing documents to be embedded in the same space as the words it contains. To do so, each document is combined with the context words to predict the target word. That is, given a document d, context words $w_{t-k}, \cdots, w_{t-1}$, and target word w_t, the goal is to learn the model $p(w_t|d, w_{t-k}, \cdots, w_{t-1})$. The same neural network is used to estimate this model as in Word2Vec. See Figure 2.1 for an illustration of this process. The authors show that Doc2Vec works better than other document representations like bag-of-words or tf-idf due to its ability to capture semantic relationships, lower-dimensional representation, and consideration of word order.

Once Doc2Vec embedding vectors are obtained for each document, one can then attempt to segment the space of documents using the usual clustering techniques. The simplest and quickest way to do so is via k-means. K-means can be useful for selecting representative documents to summarize because it allows one to use cluster centers as a way of estimating how much a particular document is representative of its cluster. We can presumably use this to select representative documents to summarize by using documents nearest to each cluster center as representative, which is what we do.

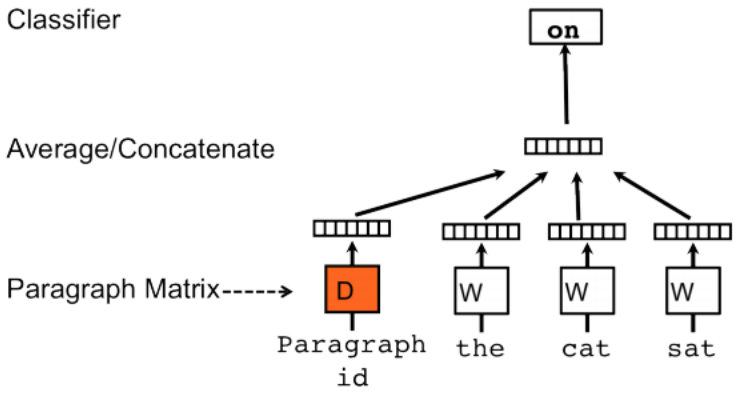

Figure 2.1: Framework for training Doc2Vec vectors. The document is combined with context words to predict a target word [1].

2.2 Sequence-to-Sequence with Attention

The first abstractive text summarization model used on our dataset is the Sequence-to-Sequence Abstractive Text Summarization model using Attentional Encoder-Decoder Recurrent Neural Networks. The model is a traditional sequence-to-sequence model with attention that is customized for text summarization. A pointer mechanism is also added to generate words not in the vocabulary. The baseline model corresponds to the neural machine translation model where the encoder consists of a bidirectional GRU, while the decoder consists of a unidirectional GRU [2].

The decoder has the same hidden-state as the encoder. It also consists of an attention mechanism over the source-hidden states and a softmax layer over the target vocabulary. The reason this is done is to generate the required words. In addition to the basic model, the large vocabulary trick (LVT) has also been adapted to abstractive summarization. LVT is a method proposed that allows us to use a very large target vocabulary without increasing training complexity. This is based on importance sampling. In this particular method proposed, decoding can be carried out efficiently, also in the case where the model has a very large target vocabulary. [7].

In this model, the most frequent words in the target dictionary are added

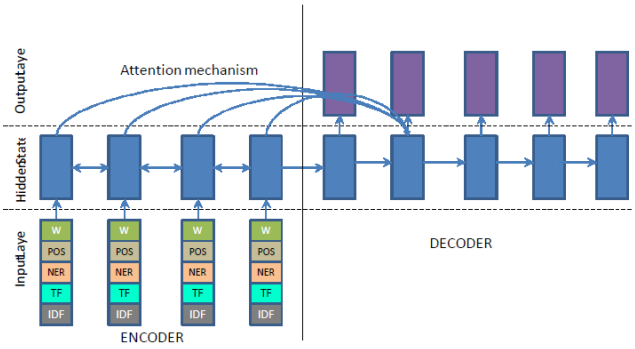

Figure 2.2: Illustration of feature-rich-encoder [2].

until the vocabulary reaches a fixed size. This results in reducing the size of the softmax layer of the decoder which is a computational bottleneck. This technique is also helpful in efficiently speeding up the convergence by focusing the modeling effort only on the words that are essential to a given example. This also complements the summarization task since a large proportion of the words in the summary come from the source document.

For the task of abstractive text summarization, one of the hurdles is to identify the key concepts and key words in the document, on which the summary is based. In order to get this done correctly, additional look up based embeddings matrices are created for the vocabulary of each tag type. Similar to the way the embeddings for words for every word in the source document are looked-up, the authors chose to look up its embeddings from all of its associated tags and concatenate them into a single long vector, as described in the figure below. Here, a feature rich encoder is used where one embedding vector is used for POS, NER tags and TF and IDF values, which are joined together with word-based embeddings that serve as an input to the encoder [2]. On the target side, the authors continued to use only word-based embeddings as the representation.

To deal with the case where the word is not in the vocabulary of the decoder, the traditional way is to emit a UNK token as a placeholder. However, the summary produced from this mechanism is usually not legible. In summa-

rization, we handle such out-of-vocabulary (OOV) words by pointing to their location in the source document [2].

In the proposed model, the decoder is equipped with a switch that decides between using the generator or a pointer at every time-step. The decoder produces a word from its target vocabulary in the normal fashion if the switch is turned on, and generates a pointer to one of the word positions in the source otherwise. The word at the word position will be copied into the summary in place of the UNK token.

The switch that decides which mechanism to choose is modeled as a sigmoid activation function based on the entire available context at each timestep as shown in the following equation. In that equation $P(s_i = 1)$ is defined to be the probability of the switch turning on at the i^{th} time-step of the decoder,$\mathbf{h}_i$ is the hidden state, $\mathbf{E}[o_{i-1}]$ is the embedding vector of the emission from the previous time step, $\mathbf{c_i}$ is the attention-weighted context vector, and $\mathbf{W}_h^s, \mathbf{W}_e^s, \mathbf{W}_c^s, \mathbf{b}^s$ and $\mathbf{v}^s$ are the switch parameters.

$$\begin{aligned} P(s_i = 1) &= \sigma(\mathbf{v}^s \cdot (\mathbf{W}_h^s \mathbf{h}_i + \mathbf{W}_e^s \mathbf{E}[o_{i-1}] \\ &+ \mathbf{W}_c^s \mathbf{c}_i + \mathbf{b}^s)), \end{aligned}$$

Attention distribution over word positions in the document is used to sample the pointer. In the following equation, p_i is the pointer value at i^{th} word-position in the summary, $P_i^a(j)$ is the probability of the i^{th} time-step in the decoder pointing to the j^{th} position in the document, and $\mathbf{h}_j^d$ is the encoder's hidden state at position j.

$$\begin{aligned} P_i^a(j) &\propto \exp(\mathbf{v}^a \cdot (\mathbf{W}_h^a \mathbf{h}_{i-1} + \mathbf{W}_e^a \mathbf{E}[o_{i-1}] \\ &+ \mathbf{W}_c^a \mathbf{h}_j^d + \mathbf{b}^a)), \\ p_i &= \arg\max_j (P_i^a(j)) \text{ for } j \in \{1, \ldots, N_d\}. \end{aligned}$$

At training time, the model is provided with explicit pointer information whenever the summary word does not exist in the target vocabulary. When

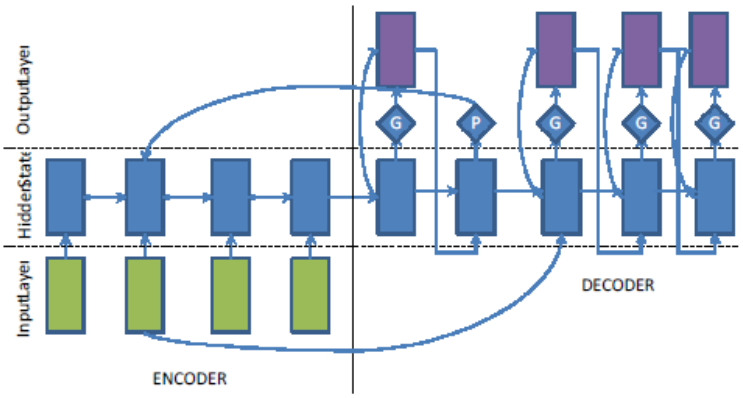

Figure 2.3: Illustration of the switching generator/pointer model [2].

the OOV word in the summary occurs in multiple document positions, the tie is broken down in favor of its first occurrence. At test time, the model decides automatically at each time-step whether to generate or to point, based on the estimated switch probability $\mathrm{P}(s_i)$. Therefore, in the above manner, this model summarizes an article in an abstractive method.

2.3 Pointer-Generator Network

Another model we apply is the Pointer-Generator Network (PGN) from [4]. In this paper, the authors use a Seq2Seq network with attention as baseline. A Seq2Seq model usually consists of three parts: the encoder, the decoder, and the attention mechanism. In this paper, the encoder part adopts a single-layer bidirectional LSTM, inputting the word vector sequence of the original text, and outputting a encoder hidden state sequence h_i. The decoder part adopts a single-layer unidirectional LSTM, and the input of each step is the word predicted in the previous step. The word vector simultaneously outputs a decoded state sequence s_t for the prediction of the current step. The attention is the probability distribution for the original text. The purpose is to tell the model which words in the original text are more important in the prediction process of the current step. The formulas are

$$e_i^t = v^t \tanh(W_h h_i + W_s s_t + b_{attn})$$
$$a^t = \mathrm{softmax}(e^t)$$

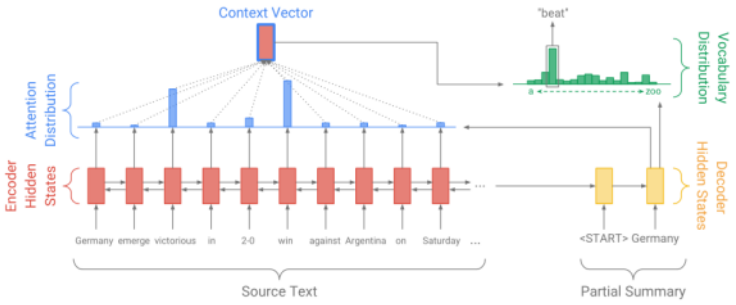

Figure 2.4: Illustration of a seq2seq model with attention.

After calculating the attention distribution of the current step, they derive a weighted average of the hidden layer of the encoder output, and have the dynamic representation of the original text, which is called the context vector.

$$h_t^* = \sum_i a_i^t h_i$$

Finally, based on the hidden layer and context vector of the decoder output, both of them determine the probability distribution on the vocabulary in the current prediction step.

$$P_{vocab} = \text{softmax}(V'(V(s_t, h_{t*}) + b) + b')$$

The loss function used is defined by

$$loss = \frac{1}{T} \sum_{t=0}^{T} - \log P(w_t^*)$$

The PGN is a hybrid model of the Seq2Seq model and the pointer network. On the one hand, it maintains the ability to generate abstractions through the Seq2Seq model. On the other hand, it directly extracts words from the original text based on the pointer network, improving the accuracy of the summary and alleviating the OOV problem. At each step of the prediction, the pointer-generator network combines two models together by dynamically calculating a generation probability p_{gen}.

$$p_{gen} = \sigma(w_h^T h_t^* + w_s^T s_t + w_x^T x_t + b_{ptr})$$

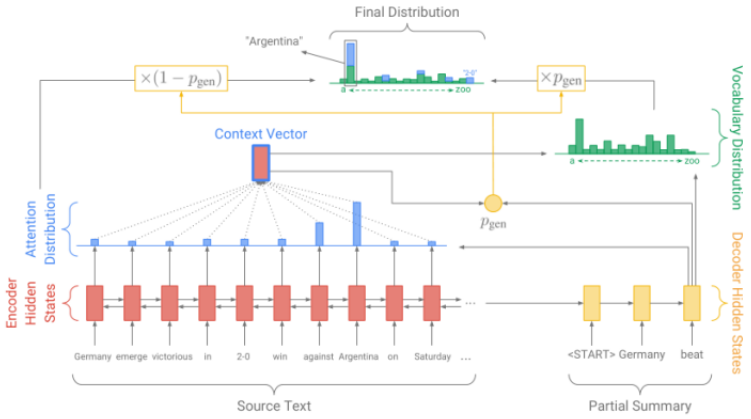

Figure 2.5: Illustration of a pointer-generator network.

This paper handles the pointer network pretty ingeniously. The attention distribution calculated by the Seq2Seq model is directly used as the output of the pointer network. By parameter multiplexing, the complexity of the model is greatly reduced. The final prediction is

$$P(w) = p_{gen}P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i} a_i^t$$

The PGN is equivalent to dynamically adding the original text to the vocabulary during each summary generation process. In each step of the prediction process, compared to the simple seq2seq model, the words appearing in the original text are more likely to be selected to build the summary.

In this model, they implement the coverage mechanism to deal with repeated content. The main point of the coverage model is to maintain a coverage vector during the prediction process.

$$c^t = \sum_{i=0}^{t-1} \lim a^i$$

This vector is the cumulative sum of the attention distributions calculated by all the prediction steps in the past. It records which words of the original text have been paid attention to by the model, and makes this coverage vector affect the attention calculation of the current step.

$$e_i^t = v^T \tanh(W_h h_i + W_s s_t + w_c c_i^t + b_{attn})$$

The purpose of this is to tell the model that it has been interested in the current step attention calculation, and hope to avoid the situation of continuous attention to certain words. At the same time, the coverage model adds an additional coverage loss to penalize duplicate attention. Noting that this loss only penalizes repeated attentions and does not force the model to focus on every word in the original text.

$$covloss_t = \sum_i \min(a_i^t, c_i^t)$$

Finally, the loss function in this model is

$$loss_t = -\log P(w_t^*) + \lambda \sum_i \min(a_i^t, c_i^t).$$

2.4 Reinforced Extractor Abstractor Network

The last model we use is the Reinforced Extractor-Abstractor Hybrid Network with policy-based reinforcement learning (REAN) from [3]. Given a document, the model first uses an extractor agent to select salient sentences that will be used for the summarization. The model then uses an abstractor network to rewrite the salient sentences, producing an abstractive summarization. The extractor agent and the abstractor network are bridged with a reinforcement learning technique, also allowing the model to be trained efficiently.

The extractor agent has two phases: sentence representation and sentence selection. During the representation phase, the agent first uses a temporal convolutional model to compute r_j, the representation of each sentence in the document. The agent then applies bidirectional LTSM on the output, optimizing the representation as h_j, taking consideration of the sentence dependency with the entire document. During the selection phase, the agent selects extracted sentences from the given representation by using another LSTM to train a Pointer Network [8]. The extraction probability is calcu-

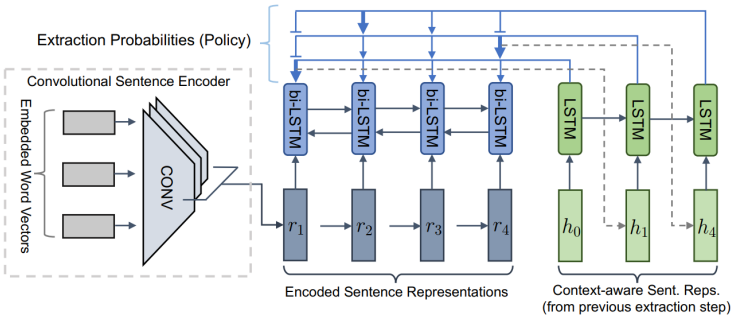

Figure 2.6: Illustration of the extractor agent [3]

lated as follows:

$$u_j^t = \begin{cases} v_p^\top \tanh(W_{p1} h_j + W_{p2} e_t) & \text{if } j_t \neq j_k \\ & \forall k < t \\ -\infty & \text{otherwise} \end{cases}$$

$$P(j_t | j_1, \ldots, j_{t-1}) = \text{softmax}(u^t)$$

where e_t's are the output of the *glimpse* operation:

$$\begin{aligned} a_j^t &= v_g^\top \tanh(W_{g1} h_j + W_{g2} z_t) \\ \alpha^t &= \text{softmax}(a^t) \\ e_t &= \sum_j \alpha_j^t W_{g1} h_j \end{aligned}$$

In the above equations, z_t is the output of the added LSTM (shown in green in Fig. 2.5) which is referred to as the decoder. At each time step t, the decoder performs a 2-hop attention mechanism: It first attends to h_j's to get a context vector e_t and then again for the extraction probabilities.

The abstractor network approximates g, which compresses and paraphrases an extracted document sentence to a concise summary sentence. The network is a standard encoder-aligner-decoder; it uses pointer-generator to include OOV words in the abstract summary. The network also uses beam-search trigram avoidance [5] at sentence level to avoid repeating and redundant

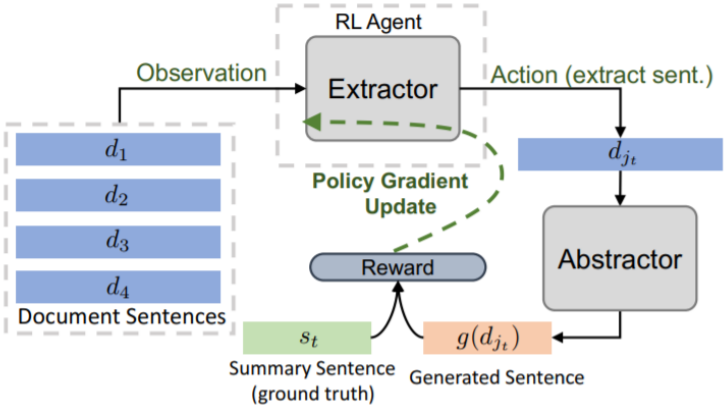

Figure 2.7: Reinforced training of the extractor (for one extraction step) and its interaction with the abstractor. [3]

words and phrases.

The model uses reinforcement learning, specifically policy gradient techniques for training optimization. The technique uses a vanilla policy gradient algorithm, REINFORCE [9], with critic network to mitigate high variance. The optimization is specifically applied to the extractor agent on its selection performance evaluated by ROUGE scores. Each time the extractor chooses a sentence with a high ROUGE score, the action is encouraged with a reward, otherwise discouraged. The technique uses a stop action policy to guide sentence selection number based on the sentence representation, allowing a dynamic number of sentence selection.

To make the extractor an RL agent, the Markov Decision Process is formulated: At each extraction step t, the agent observes the current state $c_t = (D, d_{j_{t-1}})$, samples an action $j_t \sim \pi_{\theta_a,\omega}(c_t, j) = P(j)$ to extract a document sentence and receive a reward:

$$r(t+1) = \text{ROUGE-L}_{F_1}(g(d_{j_t}), s_t)$$

The critic network is then applied with trainable parameters θ_c to predict the state-value function $V^{\pi_{\theta_a,\omega}}(c)$. The predicted value of critic $b_{\theta_c,\omega}(c)$ is called the baseline, which is then used to estimate the advantage function:

$$A^{\pi_\theta}(c, j) = Q^{\pi_{\theta_a,\omega}}(c, j) - V^{\pi_{\theta_a,\omega}}(c)$$

$A^{\pi_\theta}(c, j)$ is then maximized with the policy gradient

$$\nabla_{\theta_a,\omega} J(\theta_a, \omega) = \mathbb{E}[\nabla_{\theta_a,\omega} \log \pi_\theta(c, j) A^{\pi_\theta}(c, j)].$$

Finally, the critic is trained to minimize the square loss known as the Advantage Actor-Critic (A2C).

$$L_c(\theta_c, \omega) = (b_{\theta_c,\omega}(c_t) - R_t)^2.$$

Chapter 3

Implementation

3.1 Process

We utilized the above models to generate a corpus summary by designing and employing the following pipeline. First, we perform some type of clustering on the text corpus to decide which documents to summarize. These documents, which we refer to as representative documents, are then each passed through one of our Seq2Seq models to create short summaries of each document. We then take these separate summaries and concatenate them together to create a final summary for the corpus. An illustration of this pipeline is shown in Figure 3.1.

3.2 Results

Before we could proceed with the above pipeline we had to perform a good deal of data processing to get our collection into a format suitable for summarizing the whole collection. The data was given to us as a WARC file that was obtained by crawling the web, and moderately filtered for relevance. We first extracted the text for each archived webpage from the WARC file and filtered the text to remove noisy pages using simple rules. After filtering the text for relevance, only about 30% of the webpages were maintained and used to feed into the pipeline.

With the data cleaned, we generated corpus summaries by following the above pipeline. We used two different techniques to find representative ar-

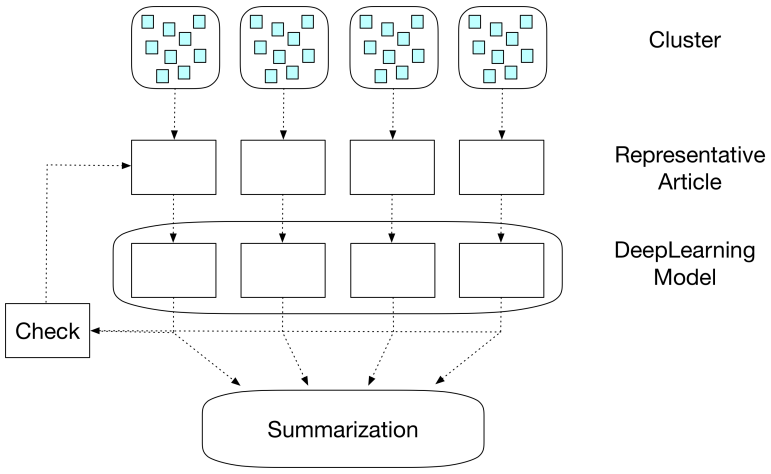

Figure 3.1: Pipeline used for text corpus summarization.

ticles to summarize. The first technique was the simplest; we randomly selected K articles from the collection and used those as our representative articles. The second technique was to perform k-means clustering on the Doc2Vec embeddings of the collection, where the K representative articles were chosen by selecting the document closest to each of the K centroids. We also attempted to use topic modeling techniques like LDA to divide the corpus into a set of topics, but we found the dataset appeared to be too homogeneous to obtain any meaningful topics.

After selecting the representative articles, we passed each article generated through the three models (Seq2Seq, PGN, REAN) described above to generate summaries for those articles. For each of the three models we decided to use pretrained models made available by the authors due to the fact that training such models on our corpus would take days to run, even on a supercomputer. Each pretrained model was trained on the CNN dataset, which is a collection of news articles scraped from cnn.com. The training summaries for the CNN dataset are obtained by using summary highlights provided by CNN for each article, which usually are no more than 1-3 sentences long. For details on the choice of hyperparameters we refer the reader to the original papers.

After obtaining summaries for each representative article, we then obtain

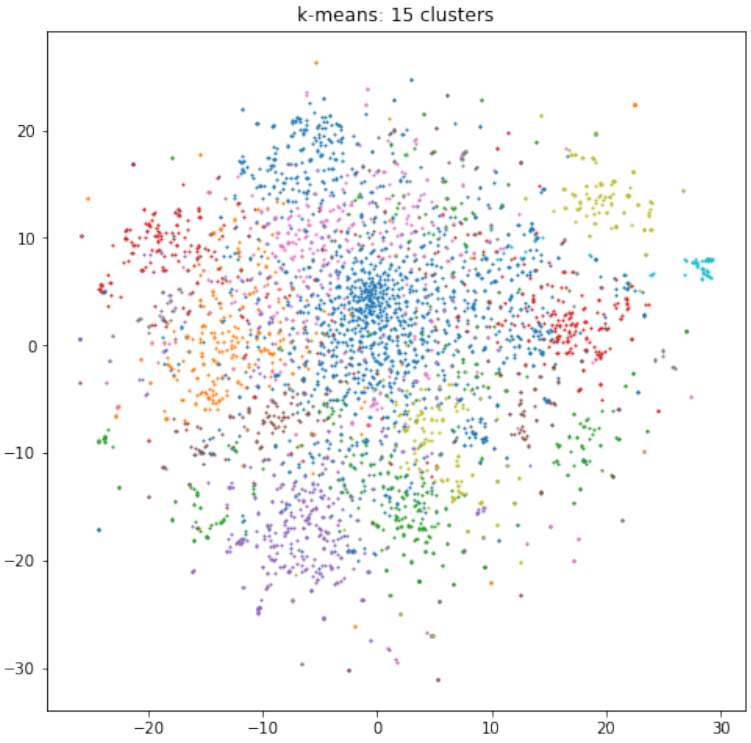

Figure 3.2: Plot of the TSNE-projected clusters obtained from performing k-means with $K = 15$ on the Doc2Vec embeddings.

our corpus summary by simply concatenating them together in the order they are selected by the above clustering. This takes our K generated short summaries, and combines them into a single long summary. Some light post-processing is also performed to make the summaries slightly more readable, including capitalizing the first letter of sentences, removing spaces before periods, and stripping out unknown tokens.

Since we explore two different ways to get representative articles (random, clustered) along with three different models (Seq2Seq, PGN, REAN), we obtained six summaries in total, one for each permutation. The value $K = 15$ was chosen experimentally and used for each case. A plot of the TSNE-projected clusters is shown in Figure 3.2. Each of the six summaries is shown in the Final Summaries section of the appendix.

Table 3.1: ROUGE scores for the generated summaries.

	ROUGE-1	ROUGE-L	ROUGE-2	ROUGE-SU4
seq2seq-random	0.0	0.0	0.0	0.0
seq2seq-clustered	0.0	0.0	0.0	0.0
pgn-random	0.152	0.091	0.094	0.049
pgn-clustered	0.030	0.030	0.0	0.005
rean-random	0.061	0.030	0.0	0.022
rean-clustered	0.091	0.091	0.031	0.022

We evaluated the quality of the final summaries by using two methods. The first method was to use an independently-created gold standard summary and do a ROUGE score comparison between the gold standard and each summary. The gold standard can be found in the Gold Standard Summaries section of the Appendix. The ROUGE scores for each summary are shown in Table 3.1. From the ROUGE scores, it appears that the PGN with random documents created the best overall summary, and the REAN with clustered documents comes next. Interestingly, the Seq2Seq model doesn't appear to generate discernible scores on either set of documents.

The next evaluation method used was a simple ranking scheme, whereby each member of the group subjectively ranked the generated summaries (1 being the best to 6 being the worst), and these rankings were averaged together to produce a subjective measure of summary quality. The rankings are shown in Table 3.2. Based on the rankings it appears the REAN model with the clustered documents created the best overall summary. Comparing with the ROUGE scores above, it's interesting to note that the PGN with clustered documents had the highest ROUGE scores, but only came 4^{th} in the subjective rankings.

Table 3.2: Rankings for the generated summaries.

	Ryan	Sudha	Chao	Li	Jiacheng	Average
seq2seq-random	6	6	6	5	6	5.8
seq2seq-clustered	5	4	5	6	5	5.0
pgn-random	4	5	4	1	3	3.4
pgn-clustered	2	2	2	3	4	2.6
rean-random	3	3	3	2	2	2.6
rean-clustered	1	1	1	4	1	1.6

Chapter 4

Conclusions

4.1 Lessons Learned

To manage this project we maintained the following division of labor: Ryan was responsible for the project management, data preparation and cleaning, Doc2Vec embedding creating, clustering, and extraction of representative articles. Sudha and Jiacheng were together responsible for generating summaries using the Seq2Seq model. Chao was responsible for using the PGN to generate summaries. Finally, Li was responsible for using the REAN to generate summaries.

The aforementioned division of project responsibilities is shown in Table 4.1. Each member also contributed a summary of lessons learned; these are shown in the comment blocks below.

Table 4.1: Division of project responsibilities.

	Responsibilities
Ryan	data preprocessing, Doc2Vec embeddings, clustering
Sudha	Seq2Seq summaries (shared)
Chao	PGN summaries
Li	REAN summaries
Jiacheng	Seq2Seq summaries (shared), Solr indexing

Ryan

To be frank, I learned that abstractive summarization is a much more difficult problem than anticipated. I assumed in the beginning that my knowledge of NLP deep learning techniques would make this exercise somewhat trivial, just plug everything into a Seq2Seq model and be done. This certainly seems not to be the case, and I can see that NLP has a long way to go to generate good abstractive summaries, especially of entire corpora. I also learned that neither topic modeling nor document embedding is a panacea. In fact, despite my best efforts I couldn't get much benefit out of using either technique to segment the document space, even though the embeddings were able to capture semantic relationships.

Sudha

A few things I learned through the project: First, I learnt that the DUC and Gigaword datasets are used to summarize news articles to a headline. I observed that the dataset chosen to train the model plays a crucial role in the type of summary generated, i.e., for news articles we need to use data sets like DUC, Gigaword, CNN Daily Mail, etc. Second, when we generate summaries just with the Seq2Seq model, it produces unknown tokens whenever it comes across an out of vocabulary word in the resultant summary. This results in illegible summaries, where important words are not represented and this is a common problem faced with this particular model. Third, I learnt that with the help of techniques like the switching decoder/pointer architecture, we can overcome the problem of encountering UNK tokens, hence resulting in legible summaries.

Chao

First of all, I touched upon and learned the basic knowledge about deep learning, and read through the relevant papers by doing the project in this course. This is one of my greatest gains. I have never done any deep learning related work before. But now, I began to have some knowledge about what it is and how it could be applied in natural language processing. I feel that this is a very interesting and challenging area of research.

Second, I am primarily responsible for recreating the pointer-generator network model and running it on our data to get the final summary. Since the author has open source code on GitHub and provided a pre-trained model, this has saved me a lot of time. However, I encountered a lot of problems when configuring how to build the environment, e.g., using the specified version of TensorFlow in a virtual environment. While running the code, I also got a clearer understanding of the whole model by going through its different inputs and outputs. Because the model needs binary input, we have to preprocess our original text. I finished it based on the code on GitHub but I went through it and learned how to do that next time if I need to.

Finally, running our own dataset to get our summary, I have a clear understanding of the application of deep learning in natural language processing. I learned a lot of concepts that I didn't know before, such as the coverage mechanism and tokenized vocabulary. Not only that but in the subsequent summary processing, I also learned many applications via the NLTK library in Python, such as grammar-checking.

Jiacheng

The first important lesson I learned was some basic ideas of neural networks. I knew nothing about neural networks before this class, but because we chose the deep learning approach, I had to read some materials to understand how the deep learning models work. In addition, I found there are many deep learning models on Github that are easy to learn and use. I learned much from these models, like using PyTorch and TensorFlow to implement the neural network, or the approaches they used to pre-process the data. Another lesson I learned is to use Cascades to train a model. I didn't realize how many resources it took to train a model. I used a seq2seq model and I found training is pretty slow in local computer and it required powerful computers so that training can finish in a reasonable time.

Li

My takeaways: Overall, I think "random" gives a better ordering and articles that are more related to the event. Also, Seq2Seq contains a lot of repeating or unstructured sentences, making it hard to read or

understand. Last, REAN seems to capture a lot more noise than PGN, making some of its sentences unreadable. PGN overall seems to be more readable than REAN for me.

4.2 Conclusion and Future Work

We believe that, among other things, we have illustrated that deep learning techniques still have a long way to go to be useful at summarizing corpora or larger bodies of text. The summaries current deep learning models generate tend to be too short, are still prone to generating meaningless phrases, and tend to focus on trivial pieces of information in a document rather than on the phrases that capture the overall document content best.

For future work many extensions are possible. Due to very tight time constraints we were only able to try a small number of possible approaches. One possible approach would be to train the deep learning models on something other than the CNN dataset; this dataset has several shortcomings for our purposes, the most important being the fact that its target summaries were created using CNN highlights, which are never more than a sentence long, and result in articles having summaries never more than 2-3 sentences. We believe training on larger documents with larger summaries should generate improved results. One possible such idea would be to train on Wikipedia, using the introductory text as the summary and the rest for the document. We are aware that such an approach was at least partially explored in another project, but have little idea how well such an approach would work for our particular use case.

Another fruitful extension would be to use some sort of text ranking approach to decide which sentences in the document to summarize. This may help deal with the tendency of the summarizers to focus on pieces of text that aren't important to the overall content of a document.

Chapter 5

Bibliography

[1] Q. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, pp. II–1188–II–1196, JMLR.org, 2014. Accessed online at `http://dl.acm.org/citation.cfm?id=3044805.3045025`. Access date November 27, 2018.

[2] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond," in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 280–290, Association for Computational Linguistics, 2016. Accessed online at `http://aclweb.org/anthology/K16-1028`. Access date November 27, 2018.

[3] M. Bansal and Y.-C. Chen, "Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting," in *ACL*, 2018. Accessed online at `https://arxiv.org/abs/1805.11080`. Access date November 27, 2018.

[4] A. See, P. J. Liu, and C. D. Manning, "Get to the Point: Summarization with Pointer-Generator Networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1073–1083, Association for Computational Linguistics, 2017. Accessed online at `http://aclweb.org/anthology/P17-1099`. Access date November 27, 2018.

[5] R. Paulus, C. Xiong, and R. Socher, "A Deep Reinforced Model for Abstractive Summarization," *CoRR*, vol. abs/1705.04304, 2017. Accessed online at `https://arxiv.org/abs/1705.04304`. Access date November 27, 2018.

[6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, (USA), pp. 3111–3119, Curran Associates Inc., 2013. Accessed online at `http://dl.acm.org/citation.cfm?id=2999792.2999959`. Access date November 27, 2018.

[7] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, "On Using Very Large Target Vocabulary for Neural Machine Translation," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1–10, Association for Computational Linguistics, 2015. Accessed online at `http://aclweb.org/anthology/P15-1001`. Access date November 27, 2018.

[8] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer Networks," in *NIPS*, 2015. Accessed online at `https://arxiv.org/abs/1506.03134`. Access date November 27, 2018.

[9] R. J. Williams, "Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning," *Machine Learning*, vol. 8, pp. 229–256, 1992.

[10] Wikipedia, "Never Again MSD," 2018. Accessed online at `https://en.wikipedia.org/wiki/Never_Again_MSD`. Accessed November 27, 2018.

[11] N. R. Safeline, "NRS' Organization of the Month — Never Again MSD — March 2018."

[12] A. Willingham, "Slacktivism is Over. The #NeverAgain Movement is About What's Next.," 2018. Accessed online at `https://www.cnn.com/2018/03/25/us/march-for-our-lives-slacktivism-trnd/index.html`. Access date November 27, 2018.

[13] A. Torres, "Timeline: How the Never Again Movement Gained Momentum After Tragedy," 2018. Accessed online at https://www.local10.com/news/parkland-school-shooting/timeline-of-never-again-movement-. Access date November 27, 2018.

Appendix A

Manuals

A.1 User Manual

The Python script `generate_summaries.py` is included in the codebase. Note that Python 3.6 was used to develop this script. Prior versions of Python 3 may also work, but the code would need to be retooled slightly to work in Python 2 since libraries unique to Python 3 are used in the script. Packages required to execute the script include the following: Gensim, NLTK, Scikit-Learn, NumPy, Pandas, JusText, PyTorch, and TensorFlow. For instructions on how to install Python 3 or the relevant packages, see the relevant documentation.

Running this script will automatically generate the entire pipeline: extract the relevant text for each article from the WARC file, perform either Doc2Vec plus k-means clustering or random sampling to find the most relevant documents, and generate the summaries for those documents using each of the deep learning models implemented.

To execute the script, run `python generate_summaries.py` in the terminal. It will take several minutes to generate the summaries. On a 3.4 GHz Intel Core i5 the script takes about 34 minutes to run, assuming the deep learning models are pretrained. Once they are generated, the summaries will be saved in the following 6 files in the `summaries` directory:

- `seq2seq_random.txt`
- `seq2seq_doc2vec.txt`

- `pgn_random.txt`
- `pgn_doc2vec.txt`
- `rean_random.txt`
- `rean_doc2vec.txt`

For example, `pgn_random.txt` will contain the summaries of the 15 randomly sampled documents using the PGN model. Applying the same script with a different WARC file or different number of clusters is also possible with only minor modifications.

We also generated Solr indexes for those wishing to query against our collection. These can be accessed by going to `http://blacklight.cs.vt.edu:8983/solr/#/~cores/big_team6`. Detailed screen dumps for how to use this can be found at the following link: `https://docs.google.com/document/d/1zOcR36jLvMwQfez6Ae0i5n-bU7VsRzff5MHNkEy4Sm4/edit?usp=sharing`.

A.2 Developer Manual

The `python generate_summaries.py` script executes the `main` function, which contains the sequence of steps necessary to generate the summaries in the report. The helper functions are all contained in `utils.py`. The essential helper functions are:

- `process_warc`: Converts WARC file to Pandas dataframe, where each row is a news article. The columns represent the URL, HTML text, and extracted plain text.
- `filter_text`: Filters the plain text from the above dataframe by removing duplicates, white space, and undesired characters from the text.
- `get_clusters`: Uses Gensim to get Doc2Vec embeddings of text and then performs K-means clustering on those embeddings. Returns the K-means model, embeddings, and tokenized text.
- `get_repr_docs`: Uses the outputs from `get_clusters` to find and return representative documents.

- `get_random_docs`: Selects representative documents by randomly sampling the collection of processed text.
- `seq2seq_summarizer`: Generates abstractive summaries for the representative documents using the Seq2Seq model.
- `pgn_summarizer`: Generates abstractive summaries for the representative documents using the PGN model.
- `rean_summarizer`: Generates abstractive summaries for the representative documents using the REAN model.

Possible extensions to the codebase could include making the script more suitable for work across both Python 2 and Python 3, ensuring that the script runs on a wide range of operating systems, and making the code more modular by allowing the user to specify the number of clusters to use or which model to use as the summarizer. An even more ambitious task would be to allow the user to pass in his own pretrained model as the summarizer. Retooling the code for multiprocessing (e.g., with Spark) may also be useful for improving runtimes.

To perform the Solr indexing, the following list of steps can be followed:

1. Copy event folder to HDFS
2. Run ArchiveSpark script
3. Copy JSON file to the home folder
4. Copy JSON file to the home folder on Solr server (via scp)
5. Copy `json_formatter.py` from `/home/fox` to the home folder on Solr
6. Covert JSON format for Solr
7. Indexing (add data to Solr core)

Appendix B

Gold Standard Summaries

B.1 Parkland Shooting Gold Standard

The first is the independently-created gold standard we used to evaluate the summaries for our given collection of the Parkland shooting news articles. This summary was written by Alex Bochel, Jun Lee, Rohit Kumar, and William Edmisten.

Gold Standard for the Parkland Shooting

On February 14, 2018, gunman Nikolas Cruz killed seventeen people at Marjory Stoneman Douglas High School in Parkland, Florida. Fourteen students and three administrators were killed in the shooting. Seventeen people were also injured non-fatally in the shooting. Cruz began the attack at 2:21 p.m. and left the premises at 2:28 p.m. He carried out the attack without intervention with police officers, despite the fact that there was an armed deputy outside within two minutes of the start of the shooting.

The seventeen victims killed in the attack include: Joaquin Oliver, Aaron Feis, Martin Duque, Meadow Pollack, Alyssa Alhadeef, Jaime Guttenberg, Alaina Petty, Cara Loughran, Nicholas Dworet, Gina Montalto, Scott Beigel, Chris Hixon, Luke Hoyer, Helana Ramsay, Alex Schater, Peter Wang, and Carmen Schentrup. Many of these victims sacrificed their lives to help their peers. Aaron Feis, a football coach and security guard at the school, shielded students with his body to protect them from bullets. Peter Wang was fatally shot while holding the door

open to allow his peers to escape the building.

The law enforcement response was complicated by several factors. For one, the emergency response calls from cell phones within the school were routed to the Coral Springs Fire Department. On the other hand, calls from parents were directed to the Broward Sheriff's Office. Another problem was that the police radio system became overloaded, forcing officers to use hand signals to communicate. Two sheriff's deputies did arrive at the scene during the shooting, but did not enter the building to intervene.

Other problems prevented internal mitigation of the attack. For instance, teachers were unable to lock their doors from within the classroom. They were forced to go outside the room with a key to lock the door. Another issue was that the gun smoke set off fire alarms in the building, causing confusion. Teachers and students were unsure whether the event was a fire, which required evacuation, or a shooting, which required seeking cover indoors.

Nikolas Cruz is currently being charged with seventeen counts of attempted murder and seventeen counts of first-degree, premeditated murder. Cruz could be sentenced to death if he is convicted. So far, he has entered a not-guilty plea in March. Cruz is also facing allegations for attacking a jail guard. Cruz was given an order to stop dragging his shoes on the floor by a guard. Upon hearing this, he became aggressive and repeatedly punched the guard in the head, and grabbed his stun gun. The stun gun was discharged in the altercation, after which the guard regained control of his weapon and punched Cruz in the face. Cruz later withdrew to a seat and was apprehended into custody.

This shooting sparked political discussion for gun control and mental health. Student led organizations as well as politicians called for gun control after the shooting. The "Never Again" movement sought to help push for these types of changes. Another topic of interest was mental health. Nikolaus Cruz showed many signs of being troubled but no serious action was ever taken to find him help. Mainly Democrats focused on the gun laws while Republicans focused on mental health.

Nikolas was adopted at a young age and never knew his birth mother. His birth mother was a drug addict and was often violent. His adoptive mother was known as being thoughtful and disciplined. Cruz took after his birth mother. He often had violent outbursts. He had an unhealthy

obsession with guns and would sometimes shoot and torture animals. Cruz even released a video in which he described how he would carry out a school shooting. These signs were not taken seriously enough to stop the shooting from occurring. His mother told the press that she indulged him in his violent video games because they would calm his mood. This also could have led to his eventual shooting.

Cruz was able to purchase his guns legally. He instantly passed a background check and was not stopped due to his mental illness. This has been a big point made when discussing the issue of gun control. Three weeks after the shooting, Florida Governor Rick Scott signed into law the Marjory Stoneman Douglas High School Public Safety Act. This act raises the age in which people can buy guns and allows judges to bar someone from purchasing a gun if they display an act of violence. The bill also allows for the voluntary arming of school officials and a school "guardian" program in which an officer is assigned to a school facility. Advocates of this bill hope that it will make the loose gun laws in Florida more strict. Currently, buying a gun does not require a permit, private gun sales do not require a background check, gun ownership does not require the owner to register their gun. With the new restrictions, there is hope to stop the occurrence of mass shootings in the United States of America.

B.2 NeverAgain Gold Standard

The second gold standard summary is the final version of the gold standard summary we created for team 7. The topic of their collection was the #NeverAgain student group, which was a consequence of the Parkland shooting collection we summarized. This summary was created using a combination of randomly selected articles from the team 7 collection along with the summary of the movement given by Wikipedia [10]. Other relevant sources are [11], [12], and [13].

Gold Standard for the NeverAgain Movement

This collection is about the NeverAgain student group. NeverAgain (also known by the Twitter hashtags #NeverAgain and #EnoughIsEnough; see also @NeverAgainMSD and neveragain.com) is a U.S. political ac-

tion committee that promotes tighter regulation of guns to prevent gun violence. The group was formed in the aftermath of a mass shooting at the Marjory Stoneman Douglas High School (MSD) in Parkland, Florida by youth involved in the theatre program. This shooting occurred on February 14, 2018, when former student Nikolas Cruz gunned down 17 students, all of whom were killed, with an AR-15 rifle, and injured 17 others.

In response to the shooting, three MSD students–Cameron Kasky, Alex Wind, and Sofie Whitney–founded NeverAgain MSD on February 15, 2018. Other Parkland students that later joined were David Hogg, Emma Gonzalez, Delaney Tarr, Alfonso Calderon, Sarah Chadwick, Jaclyn Corin, and Ryan Deitsch. Over the next few days after the shooting, the group gained over 35,000 Facebook and Twitter followers and instantly became a national movement. Also within days of the shooting, MSD students boarded buses, demanding action from lawmakers on gun control, in Florida's capital, Tallahassee. A bestselling book "#NeverAgain: A New Generation Draws the Line" appeared 19 June 2018 by David and Lauren Hogg.

Their mission is to make schools safe, and not allow one more child to be shot at school, nor one more teacher to die to save the lives of students. Their goals include for the US Congress to pass legislation addressing gun violence.

The group staged a nationwide protest against gun violence in Washington D.C. and 800 other locations on March 24, 2018. This protest, dubbed March for our Lives, was attended by millions of people from the U.S. and other countries. The group has also been present at several town hall events, with the intent to hold legislators accountable for their stance on gun control laws.

One of the primary organizations targeted by NeverAgain is the National Rifle Association (NRA), a powerful American non-profit organization that promotes gun rights and generally opposes new gun control laws. The NeverAgain movement was able to pressure several U.S. companies to sever their ties with the NRA, including the First National

Bank of Omaha, Hertz, Avis, Enterprise, Budget, MetLife, Symantec, SimpliSafe, Delta Airlines, and American Airlines.

The political reaction to the NeverAgain movement was largely partisan, with Republican groups being against gun control measures and Democratic groups being for gun control. Nevertheless, the group has managed to attain a few political victories. Most notably, in March 2018, the Florida legislature passed the Marjory Stoneman Douglas High School Public Safety Act, which involved various gun control measures, including raising the age to buy a gun in the state from 18 to 21, banning bump stocks, and denying the mentally ill the right to buy guns. Illinois worked on similar legislation. Moreover, several notable retailers that sell guns, including Walmart and Dicks Sporting Goods, have raised the age to buy guns at their stores from 18 to 21. President Donald Trump also voiced support for a nationwide ban of bump stocks.

The NeverAgain movement has also been the subject of various misinformation and conspiracy theory campaigns. Notable attacks against the group came from Ted Nugent, Alex Jones, and Rick Santorum. In one example, various fake images of Emma Gonzalez tearing up the U.S. constitution were circulated on the Internet.

The group has also received large amounts of public and financial support from various notable people and organizations. Some celebrities who have donated large sums of money to the group include George Clooney, Oprah Winfrey, Jeffrey Katzenberg, and Steven Spielberg. Former president Barack Obama also wrote a letter expressing admiration of the group's goal to end gun violence. Several major colleges, universities, organizations, and companies have also voiced support of the movement.

Appendix C

Final Summaries

Corpus Summary - Seq2Seq with Random Documents

The victims of the united states shall be flown at half-staff at the white house. The united states of america, i hereby order that the flag of the united states shall be flown at half-staff at the white house. The flag of the united states shall be flown at half-staff for the same length of time.

Car rental companies dump the nra a major car rental conglomerate which operates three prominent national brands. The rental conglomerate which operates three prominent national brands is ending its corporate relationship with the national rifle association. The rental conglomerate which operates three prominent national brands is ending its corporate relationship with the national rifle association.

Wade paid a surprise visit to students at marjory douglas high school on twitter. He is ""looking forward to being more involved in the change that they will create"" after a shooting at the school last month left 17 dead. He says it was the first full day with regular class schedules since the feb. 14 mass shooting. ""i. he says he is ""looking forward to being more involved in the change that they will create"".

12 people died in a mass shooting on feb. 14, broward sheriff scott israel. 12 people were killed in a mass shooting on feb. 14, broward sheriff scott israel. 12 people were killed in a mass shooting on feb. 14, broward sheriff scott israel.

Donald trump speaks during a meeting with the members of the national governors association in the state dining room. Trump says he would have rushed in, unarmed, if he'd been there. Donald trump says he want all teachers to take a courageous action, and a lot of the individuals.

"we stand with the students adults are supposed to take care of children not only keep them safe. He says the students adults are supposed to take care of children not only keep them safe. He says we will make this world safer for your children to go to school. No kid should be afraid to walk outside.

"it was valentines day. It was supposed to be a day of love and joy. He says it was all real. Eventually, we transitioned to the classroom closet where we stayed for about an hour and a half. He says it was multiple voices. Then, our classroom door was opened, and someone asked if anyone was there.

Douglas has won five national math championships, has the states top marching band and boasts science. Douglas has won five national math championships, has the states top marching band and boasts science and engineering programs. Douglas has won five national math championships, has the states top marching band.

A student who survived the florida shooting wants you to see how a single ar-15 bullet butchered his foot. The florida shooting will need at least 4 surgeries and a year of rehab before he can walk again. Warner bros. Entertainment inc. Are requesting this consent on behalf of their affiliates. Which will use the information under their respective privacy policies.

The stoneman douglas theatre kids many of the leaders of the leaders of the never again movement. The students are indeed actors, thrust into roles that no teen-ager should have to play. One of the less inspiring things about the movement led. The students are actually crisis actors. A warped brand of trutherism that spread after sandy hook and has since.

Nbc's eastmond, survived the shooting at stoneman douglas high school, eastmond. The gunman faces 17 counts of premeditated murder over the feb. 14 school shooting. The gunman faces 17 counts of premeditated murder over the feb. 14 school shooting. The gunman was faces 17 counts of premeditated murder over the feb. 14 school.

A senior who survived the feb. 14 shooting at north community park. The attack has been needlessly divisive. (john florida sun-sentinel via ap) douglas emma gonzalez. The attack has been needlessly divisive. (john florida sun-sentinel via ap) douglas emma gonzalez.

The school resource officer stationed suspended without pay after he learned the deputy never went into the building. The school resource officer stationed at marjory douglas was suspended without pay after he learned the deputy never went into the building. The school resource officer stationed at marjory douglas was suspended without pay after he learned the deputy never went into the building. The school resource officer stationed at marjory douglas was suspended without pay.

David hogg is a lover of film and t.v. Production. Photo by lauren newman. His friend thomas decided to spend their summer afternoon at topaz jetty, douglas high school senior david hogg. His friend thomas decided to spend their summer afternoon at topaz jetty, douglas high school senior david hogg.

The brother of the parkland school shooting suspect was arrested on a trespassing charge. The brother of the parkland school shooting suspect was arrested on a trespassing charge. The brother of the parkland school shooting suspect was arrested on a trespassing charge. The brother of the parkland school shooting suspect was arrested on a trespassing charge.

Corpus Summary - Seq2Seq with Clustered Documents

His department wont be a part of a multi-agency task force the county task force the county task force. Israel says his department wont be a part of a multi-agency task force the county task force. He says his de-

partment wont be a part of a multi-agency task force the county task force. He says his department will have an independent national organization the police executive research forum.

Runcie medina, says he knew cruz was trying to enter the building but did not confront him. He says he knew cruz was trying to enter the building but did not confront him or issue an alert. He says he knew cruz was trying to enter the building but did not confront him or issue an alert. He says he knew cruz was trying to enter the building but did not confront him or issue.

The video shows former deputy scot peterson flagging down another male staff member from the school at the main administration building. The video shows former deputy scot peterson flagging down another male staff member from the school at the main administration building. The video shows former deputy scot peterson flagging down another male staff member from the school at the main administration building. The video shows former deputy scot peterson flagging down another male staff member.

"kyle kashuv, survived the shooting at marjory stoneman douglas last month, stoneman douglas high school student. A bill to address school violence. Olivier doulierytns kyle kashuv, survived the shooting at marjory douglas high school student. Olivier doulierytns kyle kashuv, survived the shooting at marjory stoneman douglas last month, olivier doulierytns. A bill to address school violence. Olivier doulierytns a conservative parkland student helps set the agenda.

Cruz will ""get into a school and shoot the place up"": full text update:. The broward police have released the full list of 23 calls for service to the home of attacker nikolas cruz. The broward police have released the full list of 23 calls for service to the home of attacker nikolas cruz.

Two parents of victims at marjory douglas high have turned into a nationally watched race. The campaigns of lori alhadeff and ryan petty are rooted in grief, anger. The campaigns of lori alhadeff and ryan petty are rooted in grief, anger. The school district failed to properly respond to

warning signs that cruz was a troubled student.

Scott beigel wins a run at pine trails park to honor their fallen coach, scott beigel. The other victims who died during the mass shooting. (video cross country team organized a run at pine trails park. The other victims who died during the mass shooting. (video cross country team organized a run at pine trails park.

David smiley and carli teproffmiami herald roundly vilified for not entering a parkland high school. He was fleeing the building after killing 17 people, peterson warned his fellow officers to stay away even as wounded students. He says he was attacked and that he was told ""they were going to gut me"" if he returned. In a 911 call. He says he was told ""they were going to gut me"" if he returned. In a 911 call.

"a rainbow appears over parkland, florida school as teachers return for first work day. The teachers of stoneman douglas high school reported back to their classrooms almost two weeks. The teachers of stoneman douglas high school reported back to their classrooms almost two weeks.

A school resource deputy failing to enter the building where a former student killed 17 people. A school resource deputy failing to enter the building where a former student killed 17 people and wounded more than a dozen others last month. The video shows a former student killed 17 people and wounded more than a dozen others last month.

"natalie lifson is a playwright, producer, and screenwriter with a lot of opinions. Www.natalielifson.com. he says he was expelled from a year earlier, and, douglas high school shooting. He says he was a member of the rof and participated in one or more of their training drills.

"marjory douglas high has fences, gates and emergency procedures to keep students safe. The school resource officer, cruz, douglas high has fences, gates and emergency procedures. The school resource officer, cruz, douglas high has fences, gates and emergency procedures to keep students safe. The school day ended. Cruz, who had been expelled from the school for behavioral problems.

Rob schmitt says she was able to get to florida at the last minute to attend a funeral for a childhood friend. She says she was able to get to florida at the last minute to attend a funeral for a childhood friend. She says she was able to get to florida at the last minute to attend a funeral for a childhood friend. She says she was able to get to florida at the last minute to attend a funeral for a childhood friend.

A former student opened fire with a semi-automatic rifle at least 17 people and sending hundreds fleeing into the streets. The shooter, douglas high school shooting"" olivia, is being held in a broward county jail on charges of premeditated murder. The shooter, douglas high school shooting. Olivia, is being held in a broward county jail on charges of premeditated murder.

"two stoneman douglas high school junior jonathan blank who recalled the horrific day 17 people. Jonathan blank who recalled the horrific day 17 people were shot and killed at his school in parkland. Douglas high school. Jonathan blank is a junior at marjory stoneman douglas high school who survived the mass shooting. Jonathan blank is a junior at marjory douglas high school who survived the mass shooting.

Corpus Summary - PGN with Random Documents

Presidential proclamation honoring the victims of the tragedy in parkland, florida. As a mark of solemn respect for the victims of the terrible act of violence perpetrated on february 14, 2018, by the authority vested in me as president of the united states of america, i hereby order that the flag of the united states shall be flown at half-staff at the white house and upon all public buildings and grounds, at all military posts and naval stations. In witness whereof, i have hereunto set my hand this fifteenth day of february, in the year of our lord two thousand eighteen years.

Conglomerate which operates three prominent national brands is ending its corporate relationship with the national rifle association. This is the law that could take down trump it 's all well and good for paul manafort to spend 15 years in prison for money laundering, and rick gates to

plead guilty to cut a deal, and alex van der zwaan to miss the birth of his first child because he lied to the fbi. Laura moser democratic voters need to hear that laura moser is not going to change washington.

Miami heat guard dwyane wade paid a surprise visit to students at marjory stoneman douglas high school on wednesday. He is " looking forward to being more involved in the change that they will create " after a shooting at the school last month left 17 dead. " i just wanted to come and say i 'm inspired by all of you, " wade wrote in a second tweet.

The suspected shooter, identified as nikolas cruz, 19, is in custody. Cruz, a former student who was expelled last year, used an ar-15 rifle and had multiple magazines on him, according to bso. Broward sheriff scott israel said the shooter had a gas mask on, pulled the fire alarm then used smoke grenades as students ran out.

Donald trump says he would have rushed into florida school, unarmed, if he 'd been there. Trump 's heroic hypothetical vividly demonstrated his frustration at the way the deadly events unfolded at the school in parkland, florida. President donald trump says he does n't want all teachers to have guns.

After years of attacks by people with weapons of war, students can not feel safe and are demanding those weapons to be brought under control. But after years of attacks by angry, motivated students from marjory stoneman douglas high school, in parkland, fla, where 17 people were killed by a former student with an assault weapon have had the courage to take care of children not only keep them safe. They are not asking for their schools to become armed garrisons.

For the majority of the day, it was supposed to be a day of love and joy, and for the majority of the day. It was supposed to be a day of love and joy, and thats what many of us thought it was. Up until 2:20, when the fire alarm went off, and huddled into the corner of the classroom. Stoneman douglas high school is the jewel of parkland, graduating top-notch students and athletes who grow up in a dignified affluence far removed from the gritty urban sprawl of miami. Many live in vast gated com-

munities enveloped by horse pastures and pristine nature trails. Rated among the best high schools in florida, stoneman douglas has won five national math championships.

Kyle laman was inside marjory douglas high when nikolas cruz opened fire and brutally murdered 17 people. The 15-year-old freshman will need at least 4 surgeries and a year of rehab before he can walk again. Kyle 's mom, marie, used the tv show " the walking dead " as a way of describing her son 's injury.

Survivors of the marjory stoneman douglas high school shooting has been. The conspiracy theoryproliferated in the swampy depths where such [unk] the students are actually crisis actors. The spring awakening of the stoneman douglas theatre kids many of the leaders of the never again movement are indeed actors.

Stoneman eastmond says she survived the shooting at stoneman douglas high school, shares her story with nbc 's megyn kelly. The accused gunman faces 17 counts of premeditated murder over the feb. 14 school shooting. Many survivors have taken action in hopes to prevent similar school shootings from occurring.

Emma gonzalez, a senior who survived the feb. 14 shooting at marjory stoneman douglas high school, talks with people at north community park in parkland, fla, sunday, feb. 18, 2018. Gonzalez is one of the students who escaped the deadly school shooting and focused their anger at president donald trump, contending that his response to the attack has been needlessly divisive. Marjory stoneman was born in 1890 to an entrepreneur father and a concert-violinist mother.

Broward sheriff scott israel said thursday the school resource officer stationed at marjory stoneman douglas was suspended without pay after he learned the deputy never went into the building when the shooting began. Parkland school cop may not have been only officer who failed to engage shooter almost two decades ago. He may not have been the only broward deputy who didnt immediately rush into the building.

Photo by lauren newman in redondo beach, calif, marjory stoneman douglas high school senior david hogg and his friend thomas decided to spend their summer afternoon at topaz jetty, a local beach buzzing with visitors. Thomas, who was a frequent visitor to the beach, set his boogie board on a trash can to prevent it from getting sandy. As the boys returned to their spot on the sand, the lifeguard followed. Standing inches away from the young mans face in a hostile stance.

Post has learned that this article contained several passages that were largely duplicated, some without attribution, from a story by the miami herald. Post policy forbids the unattributed use of material from other sources. The bond for trespassing, kimok said, is typically $25, which zachary cruz had already posted.

Corpus Summary - PGN with Clustered Documents

Broward administrator bertha henry said the sheriffs office, which has many questions raised about its actions, is the only agency she has approached so far that has responded negatively. It is essential that we come together to collectively understand all of our efforts during this event, connect missing dots, and identify steps to close gaps in our response. The commissioners also want to create a task force. Until fdle determines the actual facts of what occurred february 14.

“ we have reviews going on and i do n’t want to get into it too much, ” runcie says. Medina, who served as a baseball coach at the school, was the first person to see shooter nikolas cruz enter school grounds on the day of the shooting. In interviews with law enforcement officials, medina later said he knew cruz was trying to enter the building but did not confront him or issue an alert which would have locked down the school.

Surveillance video shows the only armed sheriffs deputy at marjory stoneman douglas high school in parkland, fla. The 30-minute video shows former deputy scot peterson flagging down another male staff member from the school at the main administration building and hopping on a golf cart to race toward the freshman building. Mr. peterson takes cover outside the corner of an adjacent building.

Kyle kashuv, who survived the shooting at marjory stoneman douglas last month, speaks during a news conference in washington, d.c, on tuesday, march 13, 2018. He needed someone to adjust his tie, which was left in a knot so he could slip the loop around his head. In the media -lsb- trump -rsb- is portrayed as ignorant and unknowing and cold, but in real life hes.

Full text update : cruz will " get into a school and shoot the place up " the final call, however, just like the transcript from the january 5 call discussed below, shows that serious concerns were not acted on. In the interest of full transparency, we are making available the list of all 23 calls for service at the cruz home.

The campaigns of lori alhadeff and ryan petty are rooted in grief, anger and a desire to prevent another school shooting. Every parent in broward county deserves to have their child come home to them safely, petty told the sun sentinel in an interview tuesday. Theyre now supporting each others candidacy and are forming a political action committee called broward parents for better and safer schools to raise money.

Marjory stoneman douglas ' cross country team organized a run at pine trails park to honor their fallen coach, scott beigel, and the other victims who died during the mass shooting. Video originally aired 2/20/18 as thousands of runners congregated in the center of pine trails park. The latter part of the course circled around 17 makeshift crosses on the west side of the amphitheater. Regardless of the manner they ran, everyone was there to pay their respects to the attack.

Seventeen people were killed in a mass shooting at marjory stoneman douglas high school on feb. 14, 2018, in parkland, florida. David ovalle, charles rabin, david smiley and carli teproffmiami herald roundly vilified for not entering a parkland high school during a mass shooting, broward deputy scot peterson insisted publicly that he believed gunfire was happening outside on campus not from inside the building. But internal radio dispatches released by the sheriffs office thursday show peterson immediately fixated on building 12 and even radioed that gunfire was

happening inside.

Usa today a rainbow forms over marjory stoneman douglas high school on monday morning. The teachers of stoneman douglas high school reported back to their classrooms almost two weeks after the feb. 14 school shooting that claimed the lives of 17 students and teachers. Teachers will spend monday and tuesday preparing for students who return to class on wednesday. A rainbow appears over marjory stoneman douglas high school on monday.

Broward county judge jeffrey levenson ruled monday that the bso had to release the video within 24 hours of thursday 's noon deadline to appeal. Several media organizations sued the bso and broward county public schools for the release of the video recordings from the exterior of the parkland school.

Natalie lifson is a playwright, producer, and screenwriter with a lot of opinions. Douglas high school, the institution he had been expelled from a year earlier, and, according to the police report released this morning, began shooting students that he saw in the hallways and on school grounds. 17 are confirmed dead and 14 injured.

Marjory stoneman douglas high has fences, gates and emergency procedures to keep students safe. He came when he knew the gates would be open and set off a fire alarm that would dismantle a safety system, officials say. And the school resource officer, who is supposed to help protect students, may not have been on school grounds at the time.

Rob schmitt reports from parkland, florida, was one of the 17 victims who was killed in the florida high school shooting. Jordana judson, a 23-year-old now living in new york city, raced to laguardia airport thursday after hearing her friend, meadow pollack, was one of the 17 victims who was killed in last week 's shooting. Judson was able to attend the 18-year-olds funeral for a childhood friend killed in last week 's shooting.

A former student opened fire with a semi-automatic rifle at a florida high school. The shooter, who has been identified as nikolas cruz, is

being held in a broward county jail on charges of premeditated murder. It 's the 18th school shooting to take place in the united states this year. The last time i talked to nikolas cruz was my freshman year of high school.

Jonathan blank is a junior at marjory stoneman douglas high school. Jonathan was on the first floor of the freshman building in a holocaust studies class when the bullets started to fly. Many shots went off and then right away i tried to take cover, i jumped under a table and hid with two other girls under the desk.

Corpus Summary - REAN with Random Documents

Use the transcript to help students with reading comprehension and vocabulary. The flag shall be flown at half-staff for the same length of time at all united states embassies, consular offices, and other facilities and naval vessels. He set my hand this fifteenth day of february, in the year of our lord two thousand eighteen. Presidential proclamation honoring the victims of the tragedy in parkland, florida our nation grieves with those who have lost loved ones.

The nra is ending its corporate relationship with the national rifle association, as backlash grows. Trump is the law that could take down trump it's and good for paul manafort to spend 15 years in prison for money laundering, and rick gates. Trump listened quietly to entreaties for action, personal stories of grief and loss. What do jotted.

Miami heat guard dwyane wade paid a surprise visit to students at marjory stoneman douglas high school on wednesday, writing on twitter that he is "looking forward to being it was the first full day with regular class schedules since the feb. 14 mass. "i just wanted to come and say i'm inspired by all of you," wade. Students, wade arrived at the school during a lunch period and addressed a crowd in the cafeteria. Boston, ma - february 11: rodney hood #1 of the cleveland cavaliers dribbles during a game against the boston celtics at td 2018 in boston.

12 people died in the school, 2 person died. The suspected shooter, identified as nikolas cruz, 19, is in cruz was kicked out of the school.

Cruz, was expelled last year, used an ar-15 rifle and had multiple magazines on him, according to reports. Windows shattered by bullets at marjory stoneman douglas high school in parkland after 17 people were killed.

President donald trump speaks during a meeting with the national governors association in the state dining room of the white house, monday, in washington. (ap photo/evan trump says he would have rushed into florida school, president donald trump, who's. (ap photo/evan president donald trump speaks at a meeting with the members of the national governors association. The white house, trump said monday, "you don't know until you're tested, but i think most of the people in this room.

Schools are essentially an extension of the home, in that sense, providing sanctuaries of nurturing and care. After years of attacks by people with weapons of war, students feel safe and are demanding that adults end years of complaisance. Wednesdays : wednesdays walkout at schools around the country inspired by angry, motivated students. They are not asking for their schools to become armed.

It was supposed to be a day of love and joy. We followed the traditional protocol and began walking to our designated evacuation. Until 2:20, when the fire alarm went viral. Teachers yelled at us to run as well. I was in the newspaper room when it all happened.

Florida, has the states top marching band and boasts science and engineering programs where students fly weather balloons. The warning signs of a simmering danger brought on by his mental illness were documented by his fellow students, the fbi and even by his own hand, on social many live in vast gated communities enveloped by horse pastures. Nikolas cruz never felt a part of this warm nest of promise.

Kyle laman was inside marjory stoneman douglas high when nikolas cruz opened fire and brutally murdered 17 people. He will need at least 4 surgeries and a year of rehab before he can walk again. Kyle's mom, marie, used the tv show "the as a way of describing her son's. She's adamant that no one in america who isn't old enough to buy a beer should be

allowed to buy ammo.

The spring awakening of the stoneman douglas theatre kids of the never again movement are actors, thrust into roles that no teen-ager should have to be. Student. Student. " Douglas high school shooting dead. Things proliferatethat the students are actually crisis. Sex. melchior and his classmates are burning up with hormones, but in a. repressive society that insists on ignorance over truth, sex.

The gunman faces 17 counts of premeditated murder. She survived the gunman's rampage at the parkland school by using the body of one of her fellow students as a shield and by playing most of the victims were teenage students. Many survivors have taken action in hopes to prevent similar shootings.

Emma gonzalez, survived the shooting at north community park in parkland, fla, sunday, feb. 18. (john douglas emma gonzalez, survived the shooting. Gonzalez is one of the students who escaped the school shooting and focused their anger at president donald trump. Gonzalez is one of the students who escaped the deadly school shooting. Dusty hanks of bountiful and his skydiving team, sdc won their fourth consecutive national championship.

Broward sheriff scott israel said the deputy never went into the building when the shooting occurred. The school resource officer was suspended without pay. Parkland school cop may not have been only officer who failed to engage shooter almost two decades ago, when dylan klebold and eric harris went on a murderous rampage many of the three dozen killed and injured that day were struck by bullets and shrapnel long before swat. Scot peterson resigned as a broward county deputy on thursday.

Lauren newman, redondo beach, douglas high school senior david hogg and his friend thomas decided to spend their summer afternoon at topaz jetty, a local beach buzzing with as a participant in wmsd-tv, david hogg is a lover of film. Production. Through his reporting experience at wmsd-tv, hogg has proficient camera and editing skills.

Prosecutors say the article contained several passages that were largely duplicated, some without attribution, from a story. A day after the brother of the school shooting suspect was arrested on a trespassing charge at the campus of marjory stoneman douglas high school, a florida judge slapped the broward county sheriffs office also petitioned a judge to strip zachary cruz, of his right to own or possess a firearm under a new law. Cruzs attorney, joseph kimok, argued that prosecutors were being overly harsh on his client only because of his brother, nikolas cruz, who has been charged in the feb. 14 post policy forbids the unattributed use of material from other sources.

Corpus Summary - REAN with Clustered Documents

Broward sheriff's office won't be part of a multi-agency task force looking into stoneman douglas shooting incident. Event, : we come together to understand all of our efforts during this event, connect missing dots. Israel says it is not the time for the sheriffs office to be involved, since the florida department of law enforcement is doing its actions on sheriffs office, has had many questions raised about its actions, is the only agency she has responded. Israel expects the fdle investigation to take about two.

Medina, was the first person to see shooter nikolas cruz enter school grounds. Officials, medina knew cruz was trying to enter the building but did not confront him or issue an alert which would have locked down the medina was transferred from stoneman douglas after an investigation. The decision came as school leaders signed off on the job requirements for its guardianship program.

Fort lauderdale, surveillance video shows the only armed sheriffs deputy at marjory stoneman douglas high school in parkland, fla, remained outside during the feb. 14 massacre. The deputy failed to confront the gunman during the six-minute rampage, videos time stamps show. Former deputy scot peterson flagging down another male staff member from the school at the main administration building and hopping on a golf cart to race toward more than 10 minutes later, other officers appear to reach mr. petersons. The building, wounded victims lay bleeding in hallways and classrooms, according to 911 calls from students and teachers.

Olivier doulierytns survived the shooting at marjory stoneman douglas last month. Olivier doulierytns is a conservative parkland student. The high school junior who favor tighter gun-control legislation, is back in washington for a second week of high-profile meetings, and hes setting the agenda in the nations kyle kashuv, survived the shooting. He needed someone to adjust his tie, which was left in a knot so he could slip the loop.

Fbi tipster cruz will "get into a school and shoot the place up": full text update: amid public demands for transparency, the broward police. The list of all 23 calls for service at the cruz home, the sheriffs office said in a friday tweet. The most serious warning came on november 30, 2017. The final call, just like the transcript from the january 5 call discussed below, shows that serious concerns were not acted. None appeared arrestable under florida 's

The election for broward county school board has turned into a nationally watched race now that two parents of victims have filed to alhadeffs daughter, alyssa, and pettys daughter, alaina, were among the 17 who died after nikolas cruz opened fire on feb. 14. Lori alhadeff and ryan petty are rooted in grief, anger and a desire to prevent another school shooting. Theyre now supporting each others candidacy and are forming a political action committee called broward parents for better and safer schools.

Thousands participate in run honoring scott beigel and 16 other shooting victims. (video members of the marjory stoneman douglas' cross country team organized a run at pine trails park to honor their fallen coach, scott beigel, and the other victims who (video originally aired 2/20/18) as thousands of runners congregated in the center of pine trails park, members of the stoneman douglas cross country coach scott beigel. Stoneman douglas high school, a sea of runners hit the spacious pine trails fields in t-shirts and running shorts.

Seventeen people were killed in a mass shooting at marjory stoneman douglas high school on feb. 14, 2018. David smiley and carli teproffmiami

herald roundly vilified. And, cruz was fleeing the building after killing 17 people, peterson. Internal radio dispatches show peterson immediately fixated on building 12 and even radioed that gunfire was happening.

Alumni of marjory stoneman douglas high school banded together to share a powerful message with the world: "change. Teachers of stoneman douglas high school reported back to their classrooms almost two weeks after the school shooting that claimed the lives of 17 students. Usa rainbow forms over marjory douglas high school on monday morning. Teachers will spend monday and tuesday preparing for students who return to class.

Bso had to release the video within 24 hours of thursday's noon deadline. Sheriff scott israel said he failed to engage shooter nikolas cruz. Bso stoneman douglas high school surveillance video shows a school resource deputy failing to enter the building. Several media organizations sued the bso and broward county public schools for the release of the recordings from the exterior of the parkland.

The florida douglas high school shooting was an anti-semitic hate crime. Douglas high school is more than 40% jewish, cruz believed that jews were part of a conspiracy to unseat white people from power. 17 are confirmed dead and 14 have been confirmed dead. In fact, multiple of the violent incidents he was expelled for were attacks against jews.

Nikolas cruz, had been expelled from the school for 20 minutes before the school day. School resource officer, is supposed to help protect students, may not have been on school grounds. Thats the time school officials open the gates around campus so students and staff parked in various parking lots, as well as school buses. He came when he knew the gates would be open and set off a fire alarm. Marjory stoneman douglas high has fences, gates and emergency procedures to keep students safe.

Jordana judson, was one of 17 victims in the valentines day. She was able to get to florida at the last minute to attend a funeral for a childhood friend. Meadow pollack was one of the 17 victims who was killed in the florida high school shooting. (ap) judson. Parkland, '

The shooter, who has been identified as nikolas cruz, is being held in a broward county jail on charges of premeditated murder. Olivia worthington told to andrea stanley worthington on wednesday, a former student opened fire with a semi-automatic rifle at a florida high school, killing at least 17 people. By. The 18th school shooting to take place in the united states this week. "i survived the marjory stoneman douglas high school shooting" 17, 17, woke up this morning.

The horrific day 17 people were shot and killed at his school in parkland, missouri. Jonathan blank is a junior at marjory stoneman douglas high school who survived the mass shooting on feb. 14. (cbs4) jonathan was on the first floor of the freshman building in a holocaust studies class when the bullets started. I believe 4 people were shot in her.