information

Article

Change Point Detection in Terrorism-Related Online Content Using Deep Learning Derived Indicators

Ourania Theodosiadou *, Kyriaki Pantelidou, Nikolaos Bastas, Despoina Chatzakou, Theodora Tsikrika, Stefanos Vrochidis and Ioannis Kompatsiaris

Information Technologies Institute, Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece; kpantelidou@iti.gr (K.P.); nimpasta@iti.gr (N.B.); dchatzakou@iti.gr (D.C.); theodora.tsikrika@iti.gr (T.T.); stefanos@iti.gr (S.V.); ikom@iti.gr (I.K.)

* Correspondence: raniatheo@iti.gr; Tel.: +30-2311-257-793

Abstract: Given the increasing occurrence of deviant activities in online platforms, it is of paramount importance to develop methods and tools that allow in-depth analysis and understanding to then develop effective countermeasures. This work proposes a framework towards detecting statistically significant change points in terrorism-related time series, which may indicate the occurrence of events to be paid attention to. These change points may reflect changes in the attitude towards and/or engagement with terrorism-related activities and events, possibly signifying, for instance, an escalation in the radicalization process. In particular, the proposed framework involves: (i) classification of online textual data as terrorism- and hate speech-related, which can be considered as indicators of a potential criminal or terrorist activity; and (ii) change point analysis in the time series generated by these data. The use of change point detection (CPD) algorithms in the produced time series of the aforementioned indicators—either in a univariate or two-dimensional case—can lead to the estimation of statistically significant changes in their structural behavior at certain time locations. To evaluate the proposed framework, we apply it on a publicly available dataset related to jihadist forums. Finally, topic detection on the estimated change points is implemented to further assess its effectiveness.

Keywords: change point detection; terrorism; hate speech; online content; topic detection

Citation: Theodosiadou, O.; Pantelidou, K.; Bastas, N.; Chatzakou, D.; Tsikrika,T.; Vrochidis, S.; Kompatsiaris, I. Change Point Detection in Terrorism-Related Online Content Using Deep Learning Derived Indicators. *Information* **2021**, *12*, 274. https://doi.org/10.3390/info12070274

Academic Editors: Josiane Mothe and Willy Susilo

Received: 31 May 2021
Accepted: 30 June 2021
Published: 2 July 2021

1. Introduction

In recent years, considerable terrorism-related activity, including propaganda dissemination, recruitment and training, finance raising, and hate spreading towards specific social groups, has been observed in various online platforms [1]. At the same time, several advanced methods have been developed that can analyze online textual content and extract information of interest, such as affiliations towards terrorist groups and information related to terrorist events [2,3]. Such analysis can lead to the identification of key information in the fight against crime and terrorism; for instance, the early detection and analysis of crime- and terrorism-related information exchanged in online communities can promote the efficient resource allocation towards mitigating serious incidents.

The first step in this process is the detection of content of interest, and, thus far, several works have focused on developing effective classification frameworks suitable for distinguishing between terrorism vs. non-terrorism [3] or extremism vs. non-extremism content [2], among others. These methods are more oriented towards detecting suspicious content, but without focusing on the significant changes that take place over time. Such an assessment can be performed using change point detection (CPD) methods applied on suitably constructed time series which can serve as indicators of terrorism or crime activity. More specifically, one can detect significant changes in the time series of posts related to terrorism and hate speech; the position of these changes may reflect changes in the attitude towards and/or engagement with terrorism-related activities and events that trigger users

of social media platforms/forums to display a more intense online activity in the vicinity of these time points. Overall, the idea of using a CPD method in time series of terrorism- or hate speech-related posts can be seen as an alternative way to identify links between online activity and terrorism.

Towards this direction, this paper proposes a terrorism-related change point detection framework which builds on univariate and multivariate time series. Specifically, this framework facilitates the identification of points in time where statistically significant changes occur regarding the underlying data. By exploiting the temporal evolution of several indicators, such points constitute structural breaks in the behavior of the time series and may indicate the occurrence of important events where attention should be paid to. Moreover, in the case of multivariate CPD, possible correlations existing between the time series of different indicators could also be exploited.

In general, CPD methods are divided into two main categories: *online* methods [4] that aim to detect changes in real-time and *offline* methods [5] that retrospectively detect changes when considering historical data. For example, if data consisting of terrorism-related content or hate speech are considered as underlying data for the CPD algorithms, then the estimated change points based on the offline methods could offer a useful statistical analysis of such data to identify patterns and maximize the trade off between correctly identified change points and false alarms, whereas, in the case of online methods, the estimated time locations of structural breaks could enable interested parties (e.g., law enforcement) to respond in a timely manner with the aim of preventing possible radicalization, terrorist or criminal activities. In this work, our interest lies on the offline methods.

Overall, the main contribution of this work is the adoption of a change point detection method to estimate the time locations of statistically significant changes in terrorism-related time series based on a set of indicators for an effective analysis of trends and changes in a criminal context. Specifically, the detection of change points is performed in univariate as well as multivariate time series attempting to exploit possible correlations that may exist between the time series of different indicators. The presence of terrorism-related content and the expression of hate speech are detected on the basis of state-of-the-art deep learning methods (namely, Convolutional Neural Networks (CNNs)) and are used as inputs in the CPD algorithm. The evaluation carried out on data collected from a jihadist forum showcases the appropriateness of the proposed terrorism-related change point detection framework to identify changes at time locations where more attention could possibly be given. The satisfactory performance can be attributed to its ability to detect structural breaks in the time series—either univariate or multivariate—based on the time evolution of their statistical properties. To the best of our knowledge, this is the first time that change point detection algorithms are combined with the frequencies of online textual data classified as related to terrorism and/or hate speech based on well-established classification models.

The remainder of the paper is structured as follows. In Section 2, we present a brief overview of the classification and change point detection methods. In Section 3, we detail the specific setup of the proposed pipeline, whereas, in Section 4, we exhibit its applicability. In Section 5, we discuss the results. Finally, in Section 6, we summarize our main findings, argue on possible limitations of the proposed framework and provide future directions.

2. Related Work

This section reviews related work, focusing first on change point detection methods and then presenting commonly used text classification methods whose output can be the basis for effectively detecting statistically significant changes in the behavior of a time series.

Change Point Detection (CPD). Regarding the application of CPD methods in online sources (e.g., social media and Surface/Dark Web), most existing works consider Twitter data. Change point algorithms applied to time series related to Twitter posts typically aim to discover the occurrence of events of interest that could be associated with changes in the

structural behavior of the time series. For example, a nonparametric method for change point detection via density ratio estimation has been developed for tracking the degree of popularity of a given topic by monitoring the frequency of selected words [6]. Moreover, change points have been detected in Twitter streams using temporal clusters of hashtags in online conversations related to specific events [7]. CPD methods have also been combined with the outcomes of sentiment analysis in Twitter posts where the estimation of change points includes the detection of changes related to significant events [8]. Additionally, three time series produced based on tweets with positive, negative and neutral sentiment, respectively, have been used as input to change point detection towards estimating correlations among the different sentiments [9].

Concerning the use of CPD methods in terrorism-related data, the Noordin Top terrorist network data from 2001 to 2010 have been analyzed to detect significant changes in the evolution of their structure using a social network change detection method [10]. Moreover, a method for multiple change point detection in multivariate time series has been applied in a time series produced by the counts of terrorism events across twelve global regions [11]. Finally, a marked point process framework has been proposed to model the frequency and the impact of terrorist incidents based on change point analysis to search for timestamps where the process undergoes significant changes [12].

In this work, change point detection is identified as a tool to detect changes in the behavior of the time series that may indicate the occurrence of events where attention should be paid to. It is applied to terrorism-related online content, by also considering the presence of hate speech. This is achieved building upon well-established deep learning-based classification models.

Text Classification. The detection of deviant content (such as terrorism-related, extremist or abusive content) in online platforms is often addressed as a classification problem. For example, a content analysis framework has been developed in order to identify extremist-related conversations on Twitter [2]. In a similar direction, focusing on the Islamic State of Iraq and al-Sham (ISIS), content collected from social media sources has been utilized for the automatic detection of extremism propaganda [13]. Finally, a lot of effort has been placed in detecting abusive behaviors in general, such as racist and sexist content [14] or hate speech from content extracted from the white supremacist Stormfront forum [15].

Towards the development of effective classification methods, deep learning has been extensively used, with Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) methods being among the most popular ones. CNNs were originally developed to further improve image processing, resulting in groundbreaking results in recognizing objects from a pre-defined list [16]. Due to their performance in image processing tasks, CNNs gained a lot of attention and were thus subsequently applied in various Natural Language Processing (NLP) tasks, such as text classification or categorization [17], sentiment analysis [18] and machine translation [19]. In addition to CNNs, RNNs have been particularly used in NLP tasks [20]. The main difference between the two lies in the ability of RNNs to process data that come in sequences, e.g., sentences. Specifically, they analyze a text word by word and store the semantics of all the previous text in a fixed-size hidden layer [21]. Detecting terrorism-related content or the expression of hate speech in the online world can constitute an important source of knowledge for early detection of threatening situations (such as manifestation of terrorist attacks). To this end, in this work, commonly used deep learning methods are considered to develop effective text classification models, with particular focus on distinguishing between: (i) crime- and terrorism-related activities (*terrorism-related classification model*); and (ii) the expression of hate speech (*hate speech classification model*) that constitutes an indirect way of expressing violence towards a group of people (e.g., minorities). The valuable knowledge that is extracted from both the terrorist and hate speech classification models is used then as the basis of the proposed terrorism-related change point detection framework.

3. Materials and Methods

Our approach to detect statistically significant changes in content of interest, and specifically in our case in terrorism-related content, involves the following two steps: (i) classification of online material as directly related to terrorism or as containing expressions of aggressive behavior that can be considered as an initial stage which can evolve into something more dangerous (such as crime and terrorism); and (ii) change point detection that could ultimately signify the occurrence of an event of interest. Such an approach will allow the interested parties (e.g., law enforcement) to obtain a more comprehensive and thorough understanding of how crime and terrorism-related activities are carried out and evolve through time. An illustration of the overall framework is depicted in Figure 1.

Figure 1. Change point detection framework.

3.1. Classification of Online Material

First, we detail the classification framework developed for organizing content collected from online sources into two predefined sets of categories: (i) related to terrorism or not; and (ii) containing hate speech or not. As discussed, deep learning, and specifically CNNs, have gained significant popularity on NLP tasks, and therefore we opt to use them for our framework; we also experimented with RNNs without yielding any improvement in the overall performance. Specifically, two distinct CNN-based classification models are constructed, i.e., *terrorism-related classification model* and *hate speech classification model*, using the same architecture, inspired by Kim [22]; Figure 2 depicts this CNN-based model.

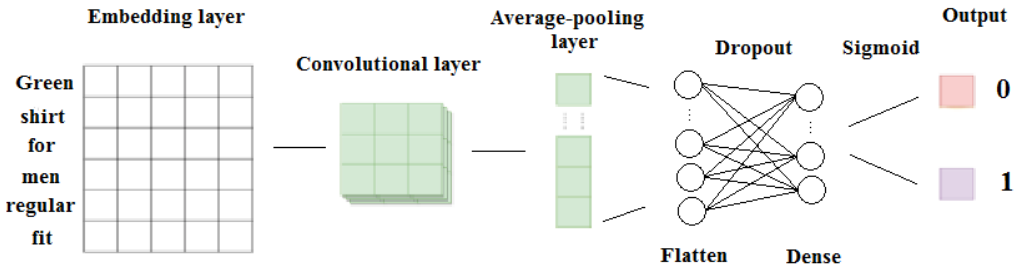

Figure 2. Overview of the CNN-based classification model.

Preprocessing. Before feeding any text to the network, a set of preprocessing steps took place to reduce noise. First, we converted the text to lowercase and then removed the hyperlinks, mentions, numbers, punctuation, accent marks, diacritics and short and long words (with <2 and >20 characters, respectively). After that, we tokenized the sequence and performed lemmatization on each term, utilizing the WordNetLemmatizer function of the nltk package (http://www.nltk.org/api/nltk.stem.html?highlight=wordnetlemmatizer; accessed on 18 March 2021).

Embedding layer. The first layer of the neural network architecture is a static embedding layer, which maps each word to a high-dimensional layer. We opted for pre-trained GloVe word embeddings to semantically represent textual content [23]. In particular, we use word vectors of dimension size 100. According to Mikolov et al. [24], 50–300 dimensions can model hundreds of millions of words with high accuracy. We experimented with word vectors of different dimensions, ranging from 50 to 200 and chose 100 due to its efficiency in terms of both performance and time of computation.

Neural Network layer. Various CNN-based architectures were tested and evaluated, by changing the number of CNN layers, filters length and kernel size. In the end, a unique

CNN layer was used, since it resulted in the best performance, with 20 filters, kernel size 3 and ReLU as activation function. A 1D average pooling layer was added on top of the convolutional layer to downsample its input, and a flatten layer followed to transform the feature map matrix into a single column. Finally, a dropout layer of $p = 0.5$ was used and sigmoid was employed as activation function. Regarding the compiling of the model, we used the Adam optimizer with learning rate 0.0001 and binary cross entropy as loss function.

To build the classification models, ground truth annotated datasets are necessary. Next, we describe the datasets used for building the terrorism and hate speech classification models.

Building the terrorism-related classification model. Due to the absence of a well-established ground truth dataset that characterizes text as terrorism or non-terrorism related, we constructed the ground truth by combining two widely used datasets: (i) the "How ISIS uses Twitter" dataset available at Kaggle (https://www.kaggle.com/fifthtribe/how-isis-uses-twitter; accessed on 26 February 2021), which contains ≈17 k tweets from 100+ pro-ISIS fanboys from all over the world since the November 2015 Paris Attacks; and (ii) the "Hate speech offensive tweets" dataset [25], which consists of ≈24 k labeled tweets organized into three classes, i.e. hate speech, offensive and neither. Since this work focuses on analyzing content in English, non-English posts were disregarded.

Overall, to build the ground truth, we considered the first dataset as terrorism-related, while the second one as non-terrorism since it is less likely to contain any terrorism-related content; the latter was constructed by randomly retrieving content from Twitter, based on a set of hate speech-related words. The newly created dataset was split into a train set of 37,973 samples, a test set of 3797 samples and a validation set of 421 samples.

Building the hate speech classification model. To build the hate speech classification model, two datasets were combined: (i) a hate speech dataset that contains texts extracted from the Stormfront [15], which consists of 1190 hate and 9462 non-hate instances; and (ii) the "Hate speech offensive tweets" [25], mentioned above, which contains ≈24 k samples categorized into three classes, i.e. hate, offensive and neither. We considered the "hate" and "offensive" instances as part of the hate class and the rest, labeled as "neither", are used for the non-hate class. Overall, the constructed ground truth dataset consists of ≈35 k samples and was split into training (90%) and test (10%) sets, maintaining the proportion of classes. From the training set, 10% was kept as validation set.

Classification Performance. To evaluate the performance of the proposed classification models, standard evaluation metrics were used, i.e., accuracy, F1-score and the Area Under Curve (AUC) value. For the terrorism-related classification model, the overall accuracy and F1-score are equal to 93%, with 99% AUC (as shown in Table 1). For the terrorism class, the model achieves F1-score equal to 91%, while the non-terrorism class obtains 94%. For the hate speech classification model, we also achieve 93% overall accuracy and F1-score (Table 2). Moreover, the AUC score is 98%. For the hate and non-hate classes, the F1-score equals 94% and 91%, respectively. Both classification models achieve particularly good performance, compared to other works that also use neural networks for text classification [26,27], which highlights the appropriateness of using them for the categorization of textual data into categories of interest.

Table 1. Performance of the terrorism-related classification model.

Accuracy	F1-Score	F1-Score Terrorism	F1-Score Non-Terrorism	AUC
0.93	0.93	0.91	0.94	0.99

Table 2. Performance of the hate speech classification model.

Accuracy	F1-Score	F1-Score Hate	F1-Score Non-Hate	AUC
0.93	0.93	0.94	0.91	0.98

3.2. Change Point Detection Method

The change point detection (CPD) method applied in this work can take into account univariate as well as multivariate time series and can be used to detect any distributional change within a sequence (e.g., regarding the mean, variance, etc.). The algorithm is called *E-Devisive* and constitutes a nonparametric approach for CPD in a set of multivariate observations [28].

Let $X_n = \{\mathbf{X}_i = (X_{i,1,}, X_{i,2}, \ldots, X_{i,d}) \in R^d, i = 1, \ldots, n\}$ and $Y_m = \{\mathbf{Y}_j = (Y_{j,1}, Y_{j,2}, \ldots, Y_{j,d}) \in R^d, j = 1, \ldots, m\}$ be independent identical distributed samples, where n and m denote the length of each sample. Samples X_n and Y_m consist of d-dimensional random variables with distributions F_1 and F_2, respectively. An empirical divergence measure is defined as follows:

$$\begin{aligned}\hat{\varepsilon}(X_n, Y_m; a) = \frac{2}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}|\mathbf{X}_i - \mathbf{Y}_j|^a \\ - \binom{n}{2}^{-1}\sum_{1\le i<k\le n}|\mathbf{X}_i - \mathbf{X}_k|^a \\ - \binom{m}{2}^{-1}\sum_{1\le j<k\le m}|\mathbf{Y}_j - \mathbf{Y}_k|^a,\end{aligned}$$

$a \in (0, 2)$. For the detection of a single change point, a scaled sample measure of the above divergence measure is defined as

$$\hat{Q}(X_n, Y_m; a) = \frac{mn}{m+n}\hat{\varepsilon}(X_n, Y_m; a), \quad a \in (0, 2).$$

Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_T \in R^d$ be an independent sequence of observations and let $1 \le \tau < \kappa \le T$ be constants, where T denotes the length of the time series of observations. The sets $X_\tau = \{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_\tau\}$ and $Y_\tau(\kappa) = \{\mathbf{Z}_{\tau+1}, \mathbf{Z}_{\tau+2}, \ldots, \mathbf{Z}_\kappa\}$ are defined, and a change point location $\hat{\tau}$ is estimated as

$$(\hat{\tau}, \hat{\kappa}) = argmax_{(\tau,\kappa)}\hat{Q}(X_\tau, Y_\tau(\kappa); a).$$

If it is known that at most one change point exists, then $\kappa = T$ is fixed.

To estimate multiple change points, the above technique is iteratively applied. Suppose that $k-1$ change points have been estimated at time locations $0 < \hat{\tau}_1 < \cdots < \hat{\tau}_{k-1} < T$. These partition the observations into k clusters $\hat{C}_1, \ldots, \hat{C}_k$, such that $\hat{C}_i = \{\mathbf{Z}_{\hat{\tau}_{i-1}+1}, \ldots, \mathbf{Z}_{\hat{\tau}_i}\}$, in which $\hat{\tau}_0 = 0$ and $\hat{\tau}_k = T$. Given these clusters, the procedure for finding a single change point is applied to the observations within each of the k clusters. The corresponding test statistic for the kth estimated change point is given by the relation $\hat{q}_k = \hat{Q}(X_{\hat{\tau}_k}, Y_{\hat{\tau}_k}(\hat{\kappa}_k); a)$, where $\hat{\tau}_k = \hat{\tau}(i)$ denotes the kth estimated change point located within cluster $\hat{C}_i$ and $\hat{\kappa}_k = \hat{\kappa}(i)$ is the corresponding constant. The running time of this iterative procedure is $\mathcal{O}(kT^2)$, where k denotes the (unknown) number of change points.

For the determination of the statistical significance (p-value) of each change point, a permutation test is implemented under the null hypothesis of no additional change points. First, the observations within each cluster are permuted to construct a new sequence of length T. Then, the estimation procedure is reapplied considering the detection of change points in the permuted observations. This process is repeated, and, after the lth permutation of the observations, the test statistic $\hat{q}_k^{(l)}$ is recorded. An approximate p-value of the kth estimated change point is defined as

$$\#\{l : \hat{q}_k^{(l)} \ge \hat{q}_k\}/(R+1),$$

where R denotes the number of random permutations.

Overall, the change point detection algorithm is implemented via the following procedure, as illustrated in Figure 3. At first, the time series is segmented into two clusters C_1, C_2 based on the time location $\hat{\tau}$ that maximizes measure $\hat{Q}$. Then, it is determined whether the estimated change point at time $\hat{\tau}$ is statistically significant or not, via a permutation test. If the estimated change point is not statistically significant, it is concluded that there

are no change points in the time series of interest. However, if the estimated change point is statistically significant, the time series is divided in two clusters of observations and the previous step is re-applied in each of these two clusters. The above-mentioned procedure is iterated in each of the clusters that is created based on the statistically significant change points that are detected, and the algorithm is terminated when no additional statistically significant change points are derived.

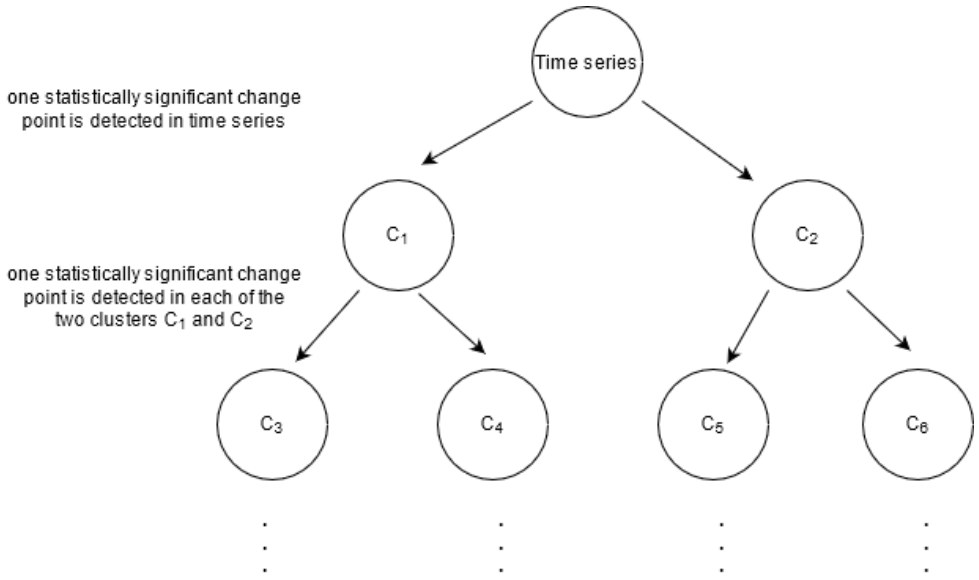

Figure 3. Overview of the CPD algorithm where $C_i, i = 1, 2, \ldots, k$ ($k - 1$ is the number of estimated change points) denote the clusters of observations that are formulated after the detection of statistically significant change points.

3.3. Dataset for Evaluation Purposes

In order to showcase the applicability of the proposed framework, we relied on the *Ansar* dataset (https://www.azsecure-data.org/dark-web-forums.html; accessed on 7 April 2021), a publicly available dataset containing terrorism-related posts. More specifically, *Ansar* is a collection of posts published in the Ansar AlJihad Network, a set of invitation-only jihadist forums in Arabic and English that are known to be popular with Western Jihadists [29]. The English portion of the dataset, referred to as Ansar1, contains 29,492 posts and spans the period 8 December 2008–20 January 2010. The dataset contains some Arabic posts, which were disregarded; after this filtering, its size equals 24,130 instances.

4. Results

This section illustrates the applicability and performance of the proposed terrorism-related change point detection framework, when applied to the Ansar1 dataset.

Extraction of indicators based on the constructed classification models. As already mentioned, both terrorism- and hate speech-related indicators are used as input to the proposed terrorism-related change point detection framework. To this end, we first exploit the classification models presented in Section 3.1, i.e., the terrorism and hate speech ones, to characterize texts as belonging to the terrorism or non-terrorism class and containing hate speech or not. The output is then exploited by the change point detection algorithm to ultimately detect previously unknown change points in the related time series that probably signify the occurrence of events of interest.

Time series. Overall, two time series are constructed and used as input to the CPD algorithm: (a) the time series of posts classified as terrorism related; and (b) those identified as containing hate speech. The posts are aggregated on a daily basis resulting in two time series with length $T = 408$ (days), which are presented in Figures 4 and 5, respectively.

These time series seem to evolve in a similar way, although the frequencies observed at the time series of the terrorism-related posts are much higher.

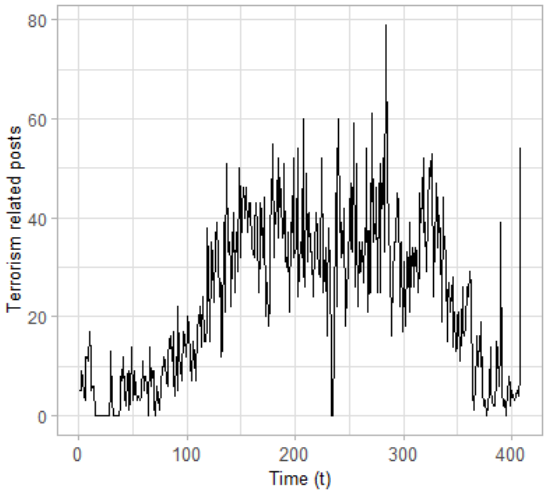

Figure 4. Time series of posts classified as terrorism related.

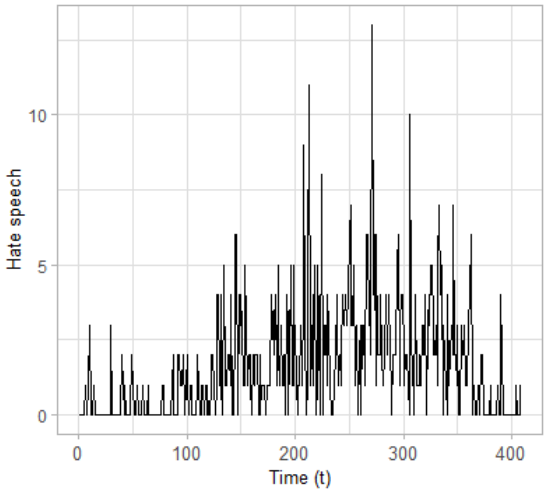

Figure 5. Time series of posts classified as containing hate speech.

Change Point Detection in the Univariate Case. The CPD method presented in Section 3.2 is applied to each of the two above mentioned time series and estimates changes in the mean value of the considered data. For the implementation of the method, we set $a = 1$ and use $R = 499$ permutations for the estimation of the statistical significance of each change point with a level of $p = 0.05$ in our significance testing. The results regarding the time series of terrorism-related posts are presented in Table 3 and graphically depicted in Figure 6a, whereas for the time series of hate speech, the results are presented in Table 4 and Figure 6b.

Considering the time series of terrorism-related posts, six change points are estimated as statistically significant (see Table 3), whereas, when the time series of posts including hate speech is considered, there are three estimated change points (see Table 4). The first estimated change point of the time series of posts related to hate speech ($t = 123$) is very close to the second estimated change point in the time series of terrorism-related posts ($t = 119$). Moreover, the third estimated change point in the time series of posts containing hate speech ($t = 363$) is close to the sixth estimated change point of the terrorism-related time series ($t = 373$).

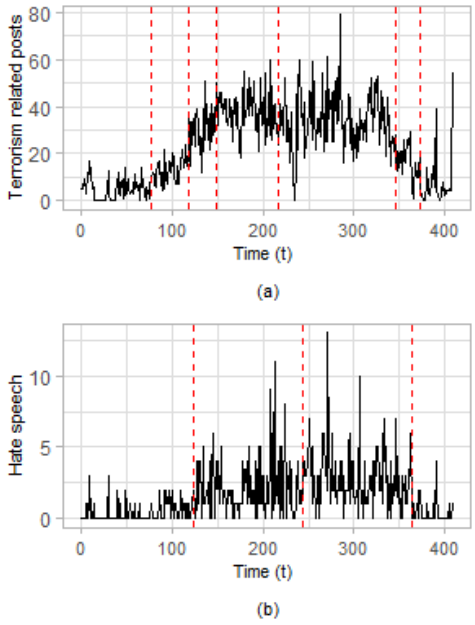

Figure 6. (**a**) Application of CPD on the time series of terrorism-related posts. (**b**) Application of CPD on the time series of posts containing hate speech; the time locations of the estimated change points are depicted as red vertical lines.

Table 3. Estimated change points for the time series with terrorism-related posts along with the corresponding significance values.

Time	Date	*p*-Value
77	23 February 2009	0.0033
119	6 April 2009	0.0033
148	5 May 2009	0.0033
216	12 July 2009	0.0033
346	19 November 2009	0.0233
373	16 December 2009	0.0465

Table 4. Estimated change points for the time series with hate speech along with the corresponding significance values.

Time	Date	*p*-Value
123	10 April 2009	0.0033
244	9 August 2009	0.0033
363	6 December 2009	0.0166

Change Point Detection in the Multivariate Case. Apart from applying CPD on the univariate case, as performed previously, we can also exploit possible correlations that may exist between the two time series using the multivariate CPD. To this end, we combine the two time series of the terrorism related posts and hate speech into a single two-dimensional time series $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_T$, $T = 408$, $\mathbf{Z}_i = (z_{i,1}, z_{i,2})$, $i = 1, 2, \ldots, 408$, where the first entry of the observation vector $\mathbf{Z}_i$ (i.e., $z_{i,1}$) is the frequency of the posts classified as terrorism-related and the second one (i.e., $z_{i,2}$) denotes the frequency of the posts classified as containing hate speech. The attempt to combine terrorism-related posts with hate speech lies on the idea that hate speech, in the sense of expressing aggressive behaviors, may be related to terrorism and vice versa. This is especially true if we consider the fact that the underlying dataset is based on jihadist forums where terrorism-related topics of discussion and the expression of aggressive behaviors may be more often. The results of the two-dimensional CPD are presented in Table 5 and depicted graphically in

Figure 7. It is observed that the estimated change points in the two-dimensional case are the same with those estimated for the univariate time series of terrorism-related posts and presented in Table 3, apart from the point at time $t = 237$. The estimated change point at time location $t = 237$ is close enough to the second estimated change point regarding the time series of posts classified as hate speech (see Table 4). Moreover, this point (i.e, $t = 237$) also appears to be a statistically significant change point for the time series of terrorism-related posts, if the value $p = 0.1$ is used for the level of our significance testing. Overall, it seems that the time series of terrorism-related posts have more impact on the two-dimensional model compared to the time series of posts related to hate speech.

Table 5. Estimated change points for the two-dimensional time series along with the corresponding significance values.

Time	Date	*p*-Value
77	23 February 2009	0.0033
119	6 April 2009	0.0033
148	5 May 2009	0.0033
216	12 July 2009	0.0033
237	2 August 2009	0.0432
346	19 November 2009	0.0133
373	16 December 2009	0.0498

Regarding the estimated change points in the two dimensional time series (Figure 7), some conclusions could be inferred the time locations of the points and the terrorist incidents that occurred during 2009 (a list of widely known terrorist attacks can be found for example: (a) at https://en.wikipedia.org/wiki/List_of_terrorist_incidents_in_2009 ; accessed on 22 April 2021, (b) at https://www.dni.gov/nctc/index.html; accessed on 22 April 2021, or (c) in [30]), which covers the main part of the *Ansar1* dataset. It can be argued that the time period between the estimated change points at time locations $t = 77$ (23 February 2009) and $t = 119$ (6 April 2009) appear to have an increasing trend, which is depicted more obviously in the frequency of posts which belong to the terrorism-related class. Therefore, the first estimated change point at $t = 77$ signals an upward change regarding the frequency of posts classified as terrorism-related and as containing hate speech, probably due to the terrorist incidents that occurred at that time.

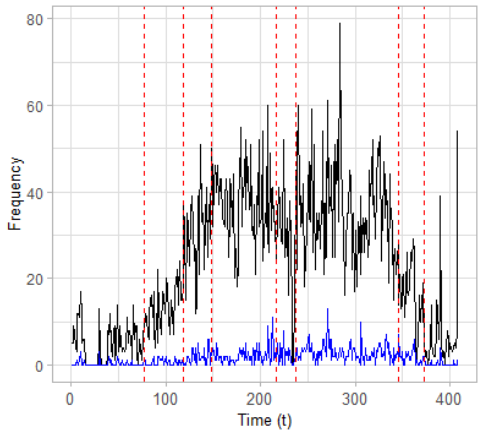

Figure 7. Application of CPD on the two-dimensional time series.

Commenting on the period that is formulated between the second estimated change point at $t = 119$ (6 April 2009) and the third one at $t = 148$ (5 May 2009), it can be argued that even more intense online activity (i.e., the trend is even more increasing) is observed compared to the previous period. This may be partially interpreted based on two factors: (a) the terrorist incidents that occurred in the previous period (e.g., Bomb explosion in

Afghanistan on 25 March 2009 and Suicide bombing in Pakistan on 27 March 2009) caused an increasing trend related to the aftermath of the attacks; and (b) other terrorist attacks took place in the period delimited by the second and third estimated change point, which enhanced the online activity. Therefore, the second estimated change point at $t = 119$ signifies an upward (and sharper) change compared to the previous period.

Regarding the period which is bounded between the third and the fourth estimated change point at time locations $t = 148$ (5 May 2009) and $t = 216$ (12 July 2009), respectively, it can be argued that the frequency of posts appears to have a stable trend at a high level compared to the previous periods. This stable trend at high frequencies may be partially explained by the two factors that are also mentioned above, i.e., the terrorist incidents that occurred in the previous period triggered an online activity that lasts and is related to the aftermaths of the attacks, and the additional terrorist incidents that occurred in the period between the third and fourth estimated change points preserved the online activity related to terrorist topics and hate speech at a high frequency level. Therefore, the third estimated change point at time $t = 148$ signals the beginning of a period with stable trend at high frequencies.

A similar interpretation of the results, as the one derived for the time period between the third and the fourth estimated change points, can also be used for the period between the fifth and sixth estimated change points at time locations $t = 237$ (2 August 2009) and $t = 346$ (19 Nvember 2009), respectively. In addition, the fourth estimated change point at time $t = 216$ signals the beginning of a short period with a decreasing trend between the two periods of stable trend at high frequencies. Finally, the two last change points estimated at time locations $t = 346$ (19 November 2009) and $t = 373$ (16 December 2009) signal the beginning of two periods with decreasing trends regarding the frequency of terrorism-related posts and hate speech, indicating partially that the interest of users among the forum has been decreased regarding terrorism-related topics.

Topic Detection. To further evaluate the effectiveness of the proposed framework, we proceed with an analysis of the topics discussed within different time periods based on the detected change points, as listed in Table 5. Specifically, we follow the Latent Dirichlet Allocation (LDA) topic detection process in each resulting time period. LDA is a generative statistical model that aims to find distinct topics in document collections [31]; to this end, it models each document as a mixture of latent topics, where a topic is described by a distribution over words. We apply the gensim version of the LDA method (https://radimrehurek.com/gensim/models/ldamodel.html; accessed on 5 May 2021). The specific parameters used for the LDA model are listed in Table 6.

Table 6. LDA parameters.

Parameter	Value
n_topics	[2–10]
random_state	100
update_every	1
chunksize	100
passes	15
alpha	'auto'
per_word_topics	True

For the topic detection, we focused mainly on the time periods where a more intense online activity is observed either via the existence of an increasing trend (6 April–5 May 2009) or via the illustration of a stable trend at a consistently high level (5 May–12 July and 2 August–19 November 2009) regarding the frequencies. For each of the aforementioned time periods, we ran the LDA method for a range of topics between 2 and 10 in steps of 1 and concluded that at most five topics resulted in a clear set of distinct topics. The results are presented in Table 7.

Table 7. Topics per time period.

	6 April 2009–5 May 2009
1	puppet, killed, vehicle, army, terrorist, destroyed, emirate, invader, province, district
2	killed, attack, police, official, soldier, force, military, people, security, troop
3	brother, group, government, sheikh, posted, pope, president, quote_originally, leader, year
4	video, case, informant, time, apostate, brother, bible, religion, corruption, john
5	militant, government, swat, security, deal, peace, district, area, region, khan
	5 May 2009–12 July 2009
1	government, force, official, militant, military, troop, region, people, country, fighting
2	brother, jihad, posted, salaam, god, people, quote_originally, video, sister, enemy
3	time, year, terrorism, arrested, religious, terrorist, woman, prisoner, human, evidence
4	killed, attack, police, soldier, source, news, killing, city, security, injured
5	killed, district, province, terrorist, invader, army, tank, emirate, puppet, destroyed
	12 July 2009–2 August 2009
1	time, situation, decision, people, political, seek, doe, salaam, land, indictment
2	god, war, religion, quote_originally, brother, posted, translated, month, apostate, video
3	police, killed, attack, soldier, wounded, officer, province, source, bomb, city
4	terrorist, killed, invader, puppet, army, tank, emirate, district, operation, province
5	force, gathering, official, suspect, intelligence, link, government, military, country, member
	2 August 2009–19 November 2009
1	sheikh, people, jihad, pour, religion, scholar, peace, son, prophet, par
2	country, group, year, war, university, government, mosque, time, link, men
3	brother, salaam, jihad, god, posted, quote, video, medium, quote_originally, protect
4	attack, killed, police, force, soldier, militant, people, official, military, security
5	terrorist, emirate, killed, district, province, area, puppet, enemy, local, time

Regarding the first time period (6 April–5 May 2009), which signals the intensification of the posting activity, we observe that the attention is highly focused on destructions and deaths related to terrorist attacks. This is in line with a set of terrorist incidents that took place in the previous period and, as a result, they may have attracted the attention of people, leading to increased online activity and intense discussions around them.

Moving on to the next time period (5 May–12 July 2009), where online activity remains at consistently high rates, there is a continuation of the discussion regarding the aftermaths of the terrorist incidents, as well as new ones that took place during this period (e.g., *20 June 2009 Taza bombing* with at least 73 deaths and more than 200 injured (https://en.wikipedia.org/wiki/2009_Taza_bombing; accessed on 22 April 2021). Now, the discussions are more oriented around the government and the military, as well as the arrests and evidence found. As expected, discussions about injuries and deaths continue with undiminished interest. Finally, there is an increased interest and discussion around issues of religion that have often been linked to terrorist attacks.

In the following short period (12 July–2 August 2009), although there is a decrease in the intensity of the discussions that take place, the attention remains on the same points with respect to the previous time period. During the last presented time period (2 August–19 November 2009), which indicates the final resurgence of interest, discussions are also beginning to focus on issues related to security, education and protection. As expected, there is insistence on discussions related to religion and god, as well as, clearly, to the deaths and killings that have occurred in the recent past.

5. Discussion

Overall, the idea of using the change point detection method in the time series of posts related to terrorism and hate speech lies on the fact that the estimation of statistically significant changes in time series at certain time positions may indicate the occurrence of events at these times that should be paid attention to. These events can be related to well

known terrorist incidents that trigger users of social media platforms/forums to illustrate a more intense online activity regarding these incidents and their aftermaths, as in our case.

Based on the results of the change point analysis regarding the retrospective detection of change points in the time series of terrorism-related posts and those containing hate speech (as presented in Section 4), some more conclusions could be inferred. At first, it can be argued that the intensity of online activity seems to be aligned with the intensity of terrorism or crime incidents to a great extent. This conclusion seems to be enhanced by the fact that, during the periods where increasing trends are depicted considering the online activity (i.e., 23 February 2009–6 April 2009 and 6 April 2009–5 May 2009) or the activity is stable at a high frequency level (i.e., 5 May 2009–12 July 2009), a considerable amount of terrorist incidents took place worldwide. Moreover, it is derived that the estimated change points associated with the increasing trends partially coincide with the time locations of terrorist incidents. This is the case for example regarding the estimated change points at times $t = 77$ (23 February 2009) and $t = 119$ (6 April 2009) where both of them signify the beginning of periods with increasing trends.

Finally, regarding the topic detection and its results, the analysis of the most popular topics discussed in the periods of greatest interest confirms the suitability of the proposed change point detection method for a better understanding of the trends around topics of interest, as well as the identification of patterns.

6. Conclusions

In this study, a change point detection framework was adopted to retrospectively detect statistically significant changes in underlying data in the context of terrorism-related activities. Specifically, a nonparametric approach was followed and applied to univariate and multivariate time series, enabling the exploitation of possible correlations that may exist between the time series of the different indicators. The proposed framework was applied on a real world dataset to display its potential in effectively detecting such changes. Both terrorism and hate speech related indicators were considered as input to the terrorism-related change point detection framework.

Based on the results of the application, it can be derived that the proposed framework could be seen as an alternative way to identify links between terrorism and online activity, since the estimated change points in the time series of frequencies are partially connected to the time locations of terrorism incidents. This implies that criminal/terrorist events trigger users of social media platforms/forums to illustrate a more intense online activity. However, depending on the forum and the users, the illustration of more intense online activity regarding terrorism or in general criminal activities may precede the occurrence of events, and in this case the proposed framework can serve as a means of early warning.

Some limitations apply in the current work. First, there is the difficulty of finding available annotated datasets, especially when focusing on the terrorism- and crime-related context. The lack of appropriate datasets in the domain prevents the comparison and cross-validation of the proposed approach in different settings. What is more, the focus on English content affects the generalization of the results. Finally, the time complexity of the CPD algorithm is quadratic in the length of the time series. In this respect, more time efficient CPD methods could be applied for very large time series.

As future work, we intend to also apply online change point detection methods in terrorism-related data, which may serve as a tool for detecting the onset in radicalization or criminal activities in real time. Moreover, additional indicators could be extracted and fed to the multivariate change point detection method, for instance, the sentiments or emotions expressed towards an event of interest.

Author Contributions: Conceptualization, O.T. and T.T.; Methodology, O.T., K.P., N.B. and D.C.; Software, O.T., K.P., N.B. and D.C. ; Writing—Original Draft Preparation, O.T. and D.C.; Writing—Review and Editing, O.T., K.P., N.B., D.C. and T.T.; Supervision, T.T., S.V. and I.K.; Project Administration, T.T., S.V. and I.K.; and Funding Acquisition, T.T., S.V. and I.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the PROPHETS (grant agreement No. 786894) and the CONNEXIONs (grant agreement No 786731) projects, funded by the European Union's Horizon 2020 research and innovation programme. The paper reflects only the authors' views and the Commission is not responsible for any use that may be made of the information it contains.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The "How ISIS uses Twitter" dataset (used for the construction of the terrorism-related classification model) is available at https://www.kaggle.com/fifthtribe/how-isis-uses-twitter; accessed on 26 February 2021. The Stormfront dataset, used for the hate speech classification model, is available at https://github.com/Vicomtech/hate-speech-dataset [15]; accessed on 26 February 2021. The "Hate speech offensive tweets" dataset, used for terrorism-related and hate speech classification models, is available at https://github.com/t-davidson/hate-speech-and-offensive-language [25]; accessed on 26 February 2021. The Ansar1 dataset, used for the validation of the approach is available at https://www.azsecure-data.org/dark-web-forums.html; accessed on 7 April 2021.

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

Abbreviations

The following abbreviations are used in this manuscript:

LDA	Latent Dirichlet Allocation
CPD	Change Point Detection
AUC	Area Under Curve
ISIS	Islamic State of Iraq and al-Sham
CNN	Convolutional Neural Networks
ReLU	Rectified Linear Unit
NLP	Natural Language Processing
RNN	Recurrent Neural Networks

References

1. Asongu, S.A.; Orim, S.M.I.; Nting, R.T. Terrorism and social media: Global evidence. *J. Glob. Inf. Technol. Manag.* **2019**, *22*, 208–228.
2. Ahmad, S.; Asghar, M.Z.; Alotaibi, F.M.; Awan, I. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Hum. Cent. Comput. Inf. Sci.* **2019**, *9*, 24.
3. Abrar, M.F.; Arefin, M.S.; Hossain, M.S. A Framework for Analyzing Real-Time Tweets to Detect Terrorist Activities. In Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 7–9 February 2019; pp. 1–6.
4. Aue, A.; Horváth, L. Structural breaks in time series. *J. Time Ser. Anal.* **2013**, *34*, 1–16.
5. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **2020**, *167*, 107299.
6. Liu, S.; Yamada, M.; Collier, N.; Sugiyama, M. Change-point detection in time-series data by relative density-ratio estimation. *Neural Netw.* **2013**, *43*, 72–83.
7. Wang, Y.; Goutte, C. Detecting changes in twitter streams using temporal clusters of hashtags. In Proceedings of the Events and Stories in the News Workshop, Vancouver, BC, Canada, 4 August 2017; pp. 10–14.
8. Tasoulis, S.K.; Vrahatis, A.G.; Georgakopoulos, S.V.; Plagianakos, V.P. Real Time Sentiment Change Detection of Twitter Data Streams. In Proceedings of the 2018 Innovations in Intelligent Systems and Applications (INISTA), Thessaloniki, Greece, 3–5 July 2018; pp. 1–6.
9. Goutte, C.; Wang, Y.; Liao, F.; Zanussi, Z.; Larkin, S.; Grinberg, Y. Eurogames16: Evaluating change detection in online conversation. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018.
10. Everton, S.F.; Cunningham, D. Detecting significant changes in dark networks. *Behav. Sci. Terror. Political Aggress.* **2013**, *5*, 94–114.
11. Tickle, S.; Eckley, I.; Fearnhead, P. A computationally efficient, high-dimensional multiple changepoint procedure with application to global terrorism incidence. *arXiv* **2020**, arXiv:2011.03599. Available online: https://arxiv.org/abs/2011.03599 (accessed on ~~30 June~~ 3 February 2021).
12. Porter, M.D.; White, G.; Mazerolle, L. Innovative methods for terrorism and counterterrorism data. In *Evidence-Based Counterterrorism Policy*; Springer: New York, NY, USA, 2012; pp. 91–112.

13. Nizzoli, L.; Avvenuti, M.; Cresci, S.; Tesconi, M. Extremist propaganda tweet classification with deep learning in realistic scenarios. In Proceedings of the 10th ACM Conference on Web Science, Boston, MA, USA, 30 June–3 July 2019; pp. 203–204.
14. Badjatiya, P.; Gupta, S.; Gupta, M.; Varma, V. Deep Learning for Hate Speech Detection in Tweets. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; pp. 759–760.
15. de Gibert, O.; Perez, N.; García-Pablos, A.; Cuadros, M. Hate Speech Dataset from a White Supremacy Forum. In Proceedings of the 2nd Workshop on Abusive Language Online (ALW2), Brussels, Belgium, 31 October 2018; pp. 11–20.
16. Guo, T.; Dong, J.; Li, H.; Gao, Y. Simple convolutional neural network on image classification. In Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China 10–12 March 2017; pp. 721–724.
17. Gambäck, B.; Sikdar, U.K. Using convolutional neural networks to classify hate-speech. In Proceedings of the First Workshop on Abusive Language Online, Vancouver, BC, Canada, 4 August 2017; pp. 85–90.
18. Tripathi, S.; Acharya, S.; Sharma, R.D.; Mittal, S.; Bhattacharya, S. Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4746–4752.
19. Gehring, J.; Auli, M.; Grangier, D.; Dauphin, Y. A Convolutional Encoder Model for Neural Machine Translation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 123–135.
20. Tamchyna, A.; Veselovská, K. UFAL at SemEval-2016 Task 5: Recurrent Neural Networks for Sentence Classification. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, CA, USA, 16–17 June 2016; pp. 367–371.
21. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent Convolutional Neural Networks for Text Classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 2267–2273.
22. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
23. Pennington, J.; Socher, R.; Manning, C. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
24. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781. Available online: https://arxiv.org/abs/1301.3781 (accessed on 3 February 2021).
25. Davidson, T.; Warmsley, D.; Macy, M.; Weber, I. Automated hate speech detection and the problem of offensive language. In Proceedings of the International AAAI Conference on Web and Social Media, Montréal, QC, Canada, 15–18 May 2017.
26. Du, J.; Gui, L.; Xu, R.; He, Y. A Convolutional Attention Model for Text Classification. In *Natural Language Processing and Chinese Computing*; Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 183–195.
27. Song, P.; Geng, C.; Li, Z. Research on Text Classification Based on Convolutional Neural Network. In Proceedings of the 2019 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 27–29 September 2019; pp. 229–232.
28. Matteson, D.S.; James, N.A. A nonparametric approach for multiple change point analysis of multivariate data. *J. Am. Stat. Assoc.* **2014**, *109*, 334–345.
29. Scanlon, J.R.; Gerber, M.S. Automatic detection of cyber-recruitment by violent extremists. *Secur. Inform.* **2014**, *3*, 1–10.
30. Burke, R.A. *Counter-Terrorism for Emergency Responders*; CRC Press: Boca Raton, FL, USA, 2017.
31. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.