Pervasive and Mobile Computing 58 (2019) 101018

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Analyzing The Emotions of Crowd For Improving The Emergency Response Services

Neha Singh *, Nirmalya Roy, Aryya Gangopadhyay

Department of Information Systems, University of Maryland Baltimore County, United States

A R T I C L E I N F O

Article history:
Received 16 October 2018
Received in revised form 15 April 2019
Accepted 20 April 2019
Available online 26 April 2019

Keywords:
Twitter
Sentiment analysis
Emotion detection
Emergency services
Change point detection

A B S T R A C T

Twitter is an extremely popular micro-blogging social platform with millions of users, generating thousands of tweets per second. The huge amount of Twitter data inspire the researchers to explore the trending topics, event detection and event tracking which help to postulate the fine-grained details and situation awareness. Obtaining situational awareness of any event is crucial in various application domains such as natural calamities, man-made disaster and emergency responses. In this paper, we advocate that data analytics on Twitter feeds can help improve the planning and rescue operations and services as provided by the emergency personnel in the event of unusual circumstances. We take an emotional change detection approach and focus on the users' emotions, concerns and feelings expressed in tweets during the emergency situations, and analyze those feelings and perceptions in the community involved during the events to provide appropriate feedback to emergency responders and local authorities. We employ improved emotion analysis and change point detection techniques to process, discover and infer the spatiotemporal sentiments of the users. We analyze the tweets from recent Las Vegas shooting (Oct. 2017) and note that the changes in the polarity of the sentiments and articulation of the emotional expressions, if captured successfully can be employed as an informative tool for providing feedback to EMS.

1. Introduction

Twitter has been there for more than a decade as a medium that allow millions of users to utilize this social platform worldwide to share the information, opinion, and broadcasting the short messages [1]. Twitter users are also known as social sensors [2,3], since they provide the information about the event as soon as it happens and eventually the discussion starts percolating on as the event unfolds. There are many applications where twitter data has been leveraged successfully in a collaborative environment [4,5]. Twitter provides the gigantic amount of data about any topic that made it useful tool for the researchers to exploit this data to identify various task such as event detection [6], event tracking [7,8], event monitoring [9] etc. Sentiment Analysis (SA) [10] is another domain which primarily focuses from the users' point of view and their sentiments about the various aspects of the event, services, product or any topic. The Sentiment Analysis summarizes the users' emotions and opinions about different topics and gives rise an aggregated idea about the user's perception and insights which is not covered in event oriented analysis or event driven task. Sentiment Analysis and Opinion Mining is a growing field of research that uses Natural Language Processing [11], Information Retrieval [12], text

* Corresponding author.
E-mail addresses: nehasingh1@umbc.edu (N. Singh), nroy@umbc.edu (N. Roy), gangopad@umbc.edu (A. Gangopadhyay).

https://doi.org/10.1016/j.pmcj.2019.04.009

analysis [13] and computational linguistics [14] to extract user's opinions and emotions towards any topics or event [15–18]. Initially, sentiment analysis was used as the voice of consumers that can identify the customer's sentiments towards any services or product. This analysis was extremely helpful for the businesses and organizations because it can provide the constructive feedback to the businesses, brands and organizations so that they can improve their services. Recently, sentiment analysis has also been applied during emergency events that deals with the sentiment's polarity (positive, negative or neutral) and various emotions of the users during the emergency event [19]. The critical details about the emotions and concerns of the people towards the crisis or emergency event are significantly important for the emergency responders to consider and reflect for improving and better managing their services to the community. Humans are driven by the sentiments most of the time, therefore, the current or future help tuned to that perception of the public will be very effective. For example, if people are fearful and angry because of some events, they are most likely not ready to listen indeed if local authorities and emergency responders have better idea about the general concern and panic of the people, it will be much easier to manage the crowd. Motivated by this, in this paper we take a bold step to combine event detection after sentiment analysis on twitter to mine the fine-grained perception of the crowd which can be piggybacked to the emergency personnel in the form of reviews of their services provided in critical situation. We investigate change point detection algorithm to detect the changing trend in the emotions and identify the associated tweets and events that caused the change in the emotion in order to send the feedback to the first responders to deal with the situation more emphatically. The main contributions of this paper can be summarized as follows:

- We collect the tweets from Las Vegas shooting event and process this dataset for inferring the sentiments and emotions of the crowd.
- We localize the various events with topic model to categorize them in their respective situational awareness.
- We propose to improve the quality of the emotion analysis in this work by enriching the existing lexicon. Expanding the emotion lexicon by combining it with Glove word embedding and retrofitting methods would enrich the specific lexicon.
- We extracted eight different emotions using NRC Emotion lexicon (EmoLex) and perform time series analysis for event detection.
- We propose a Bayesian change point detection technique to identify the changes and trends in emotions over time to provide real-time feedback to the local authorities.

Section 2 of the paper discusses the related work ans Section 3 presents the data Preprocessing methodology associating with sentiment analysis and emotion detection on the Las Vegas shooting event. Section 4 describes the overall process of the methodology. Section 5 discusses about the results and analysis, and Section 6 articulates the challenges and limitation of this work. Section 7 concludes the work and briefly explains the future goals of this research.

2. Related work

Sentiment analysis (SA) and opinion mining are growing field of research that use Natural Language Processing (NLP) and Information Retrieval (IR) to extract the users' comments, reviews, emotions and opinion towards the product, topic or event [20,21]. The recent trends in SA, and its extended application to other areas like transfer learning, emotion detection, building resources have been mentioned in [21]. Classification process can be performed mainly on three levels such as aspect levels, sentence levels and document levels. Various machine learning and Lexicon-based approaches have been used for classification of the polarity of sentiments with Naive Bayes, SVM and WordNet being the most common among all.

2.1. Sentiment analysis for disaster management

Some of the recent work has focused on identifying individual's sentiments polarity during natural disaster [19]. It has analyzed the tweets during Hurricane Sandy and projected the sentiments of users on a geographical map. The analysis shows that user's sentiments change based on their location and distance from the disaster events which may be helpful for emergency responders to get stronger situational awareness. [22] applied sentiment analysis on tweets of Genoa flood happened in October 2014 and attempted to identify the useful tweets that can provide information from the disaster management perspective. Subjective, negative, positive or ironic were the main labels for the sentiments in their dataset. The research shows that negative sentiments were more insightful than positive ones and there are specific situations that could be easily missed in complete dataset. [23] explores the usage of sentiment mining and exploits the social media in disaster management to detect people reaction during crisis situation enabling the better situation awareness. SentiStory [24] proposes a system that summarizes the events after eliminating the redundancy in microblogs based on course-grained sentiment analysis and detecting the significant changes and causes of changes in the microblogs. The study in [25] presents a visualization technique for people's emotional state during different disaster through tweet analysis. [26] research has shown that change point detection is useful in identifying potential sub-events that causes the change in sentiments of public over Twitter using the Delhi Election 2015 tweets data. The authors have used the average sentiment score (positive and negative) and applied the change point detection algorithm to see the significant changes but in this research we are focus more on identifying the temporal changes in every emotion. [19] has used the sentiment labels as positive, negative and neutral along with geo-tagged tweets but in this work we are enhancing the situation awareness by adding more fine-grained contexts through additional emotions such as anger, joy, sadness etc. SentiStory [24] posited fine-grained emotions but we propose to extract the eight different kind of emotions and capture the changes with the emotional context.

2.2. Sentiment analysis and change point detection

The combination of sentiment analysis and change point detection has the potential to uncover a lot of actionable insights if used carefully and our work is a step forward in this direction. A recent review on change point detection algorithm [27], describes its utility from ages to detect the change points in time series problems. Although, relatively less work has been explored in the area of emotional change detection along with time from social media text but there exist some work as [26], which investigates only the positive and negative sentiments of the public and we propose to look at the finer categorization of the people sentiments such as anger, anticipation, disgust, fear, joy, sadness, surprise and trust. We posit that these micro-level emotion type detection from the crowd's aggregated social footprints are insightful to design future EMS services for the population in accessing and addressing the event situation with better emotional understanding. With the best of our knowledge, this work is one of the early study which explore the time series event detection through the change in eight different user's emotions through tweets during and after of the event.

2.3. Improving the emotion analysis using external resources

Generally Sentiments polarity (positive, negative) gets considerable amount of attention in contrast of fine emotions. Creation and annotation of large scale emotion lexicon takes huge efforts and time thus, various methods that can help in the expansion of the existing lexicons would be of great interest. In this work, we have utilized the NRC EmoLex lexicon, which contains limited vocabulary (40,000 words and senses) with their associations to extract the emotions from our data. This limited vocabulary of words (lexicon) can be expanded to accommodate various other context and adapt to variety of problems such as disaster, shooting, parade, holidays etc. Since this analysis has been done on the twitter data which includes all kind of social media vocabulary including certain syntax and abbreviation terms, its important to learn the social media contextual vocabulary in order to enhance the opinion analysis with existing knowledge. There are some recent work in the area of domain adaptation in sentiment lexicon with semantic context for the social and web [28]. Transfer Learning for sentiment analysis [29] and deep sentiment analysis [30] which shows the improvement by employing the existing knowledge and adapting to the new problem lead towards better results. [31,32] refines the word embedding (Word2Vec, Glove) for sentiment analysis in order to keep the semantically and sentimentally related word close to each other. EmoWordNet [33] shows the automatic expansion of the emotion lexicon using WordNet, [34] describes the sentiment lexicon expansion based on neural PU Learning, double dictionary look-up, and polarity association. The lexicon expansion solves various problem which occurs due to the limited size of the emotion word vocabulary and lack of generalization. This issue in this work have been addressed by enriching the word vectors with lexicon knowledge and expanding the size of the lexicon using the retrofitting algorithm mentioned in [35].

3. Data preprocessing for sentiment analysis approach

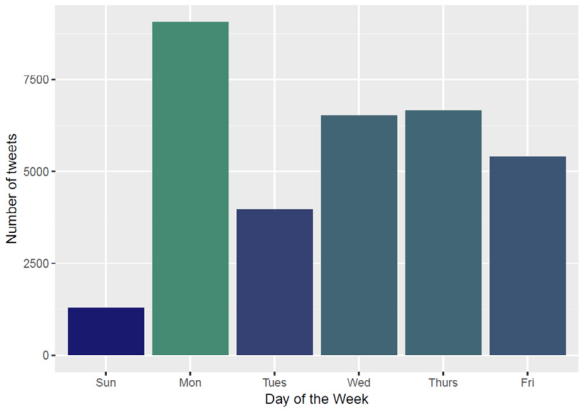
In this section, we describe the necessary steps for processing the Las Vegas shooting dataset for sentiment analysis.

3.1. Data description and data collection

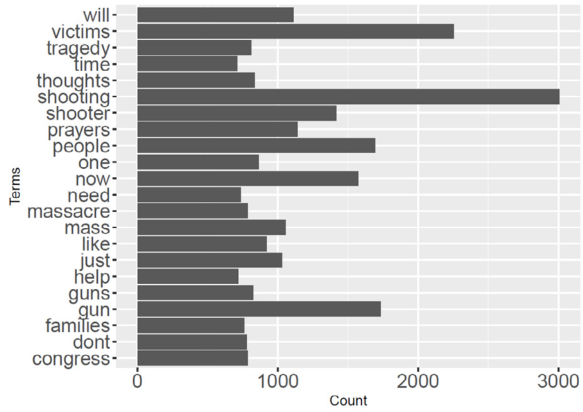
The data set used for this study is collected from Twitter right after the shooting happened in Las Vegas on October 1st, 2017 using the Twitter REST API. The data was collected using the keywords such as "*#LasVegas*", "*#LasVegas*" and "*#LasVegasShooting*" from October 1st, 2017 till October 6th, 2017. This tragic event left many lives lost and hundreds of people injured. This event quickly got the attention of people and Twitter users shared their opinion and emotions about this horrendous event. The data that we collected has total of 119,487 tweets and after removing the duplicate tweets we left with 34,859 tweets. The sentiment analysis is performed on the emotions of the users but the data also include some spam and advertisement tweets on a regular interval about their business and services which is not the target of this analysis. Thus, we removed those tweets posted by those twitter handles. Finally, 32,909 tweets were taken for further analysis along with other data fields such as tweet id, created, retweet, retweetCount etc.

3.2. Data preprocessing

The data is collected using R package "*twitteR*" and it is analyzed with the help of "*tm*" package in R. The shooting event happened on October 1, 2017 and the number of tweet distribution per day can be seen in Fig. 1(a), the primary motivation of this analysis is to understand the emotions of the people during and after the event happened. The number of tweets varies according to the intensity of the event and the time of the day. In general afternoon and late night time have the highest intensity in number of tweets. Furthermore, there are some steps taken for cleaning of the data which we depict next.



(a) Tweets Count Per Day



(b) Highly Frequent Words

Fig. 1. Data description.

3.3. Data cleaning and visualization

The data collected and saved in csv format which contains tweets along with other attributes such as created date, retweet count, screen name etc. In data cleaning all the words present in tweets are transformed into lower letters, remove the white space, special characters, url's and stopwords. Fig. 1(b) shows the frequent terms with their respective frequency and the top five frequent words are "*shooting*", "*victims*", "*gun*", "*people*". The words "*LasVegas*", "*US*" and "*vegas*" were removed from the data, since the data was collected using these keywords. Data cleaning is very important for the visualization purposes but in sentiment analysis some aspects like hashtags, punctuation, emoji's etc., are crucial and gives a lot of information in correctly identifying the emotions, thus we keep them according to the application requirements.

4. Overall process

In this section, we describe the step by step process of our twitter data analysis for sentiment categorization, and feedback mapping. Fig. 2 depicts the entire process. First, it starts with collecting twitter data and cleaning it for better visualization and understanding the characteristics of the data. Second, data is cleaned and processed in the required form where it can be analyzed further. Third, Sentiment analysis has been employed in order to extract emotions out of data that identify eight different emotions where perceptions of the victims are of utmost importance. The Emotion Detection module of this work has been expanded to improve the quality of the emotion detection. The Fig. 3 shows the extended framework for emotion detection and how a lexicon is expanded using retrofitted method to enrich the lexicon. Fourth, Bayesian change point detection algorithm is used to identify the changes in the emotions. Finally, these identified changes points represent the combination of tweets that summarize the cause of the change in emotion thus reporting it to first responders to help them better understand the general public sentiments during the crisis.

4.1. Topic modeling for event detection

Topic modeling is used for discovering the topics or main theme in a collection of documents. Latent Dirichlet Allocation (LDA) is a generative probabilistic model used for topic modeling for classifying text in any document to a specific topic [36]. Although, LDA has been used for text documents but it can be applied to various kinds of different data [37]. The working process of the LDA is explained below in Algorithm 1:

In a corpus D is a collection of M document denoted by $D = W_1, W_2, \ldots, W_M$ and each document W_m has N words denoted by $W = (w_1, w_2, \ldots, w_N)$, where w_n is the nth word in the sequence. LDA tries to find a probabilistic model of a corpus that assigns high probability to similar document and members of the corpus as well. The documents in the corpus are represented as random mixtures over latent topics and every topic is identified as a distribution over words. LDA make these following assumptions for its generative process for each document d in a corpus D:

1. Choose $N \sim Poisson(\xi)$. Poisson assumption for document length distributions
2. Choose $\theta \sim Dir(\alpha)$. Dir(α) is a Dirichlet distribution with a corpus level parameter α
3. For each of the N words w_n:
(a) Choose a topic $z_n \sim Multinomial(\theta)$
(b) Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n

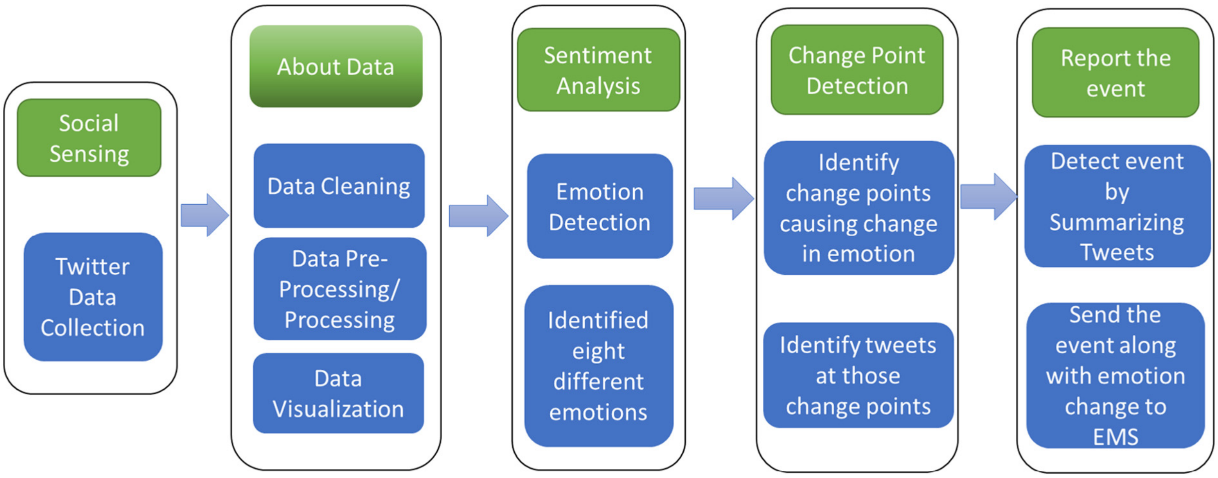


Fig. 2. Process flow chart.

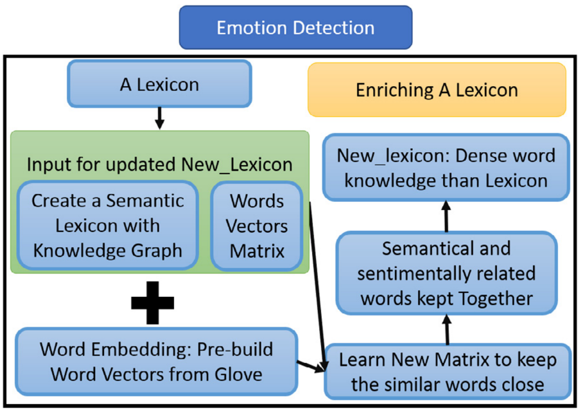


Fig. 3. Extension to improve the emotion detection.

The joint distribution of a topic mixture θ, a set of N topics z, and a set of N words w for the parameters α and β, is given by [36] $p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$

Algorithm 1 LDA for each day

procedure PERDAYLDA(Inputs: Raw Tweets with date and time ($T = (t_1, t_2, ...t_n)$),
Output: Every day k topics and terms)
Divide Tweets T into Day 1 (T_{day1}), Day 2 (T_{day2}) and Day 3 (T_{day3}) tweets
$\forall T \in T_{day1}, T_{day2}$ & T_{day3}
for T **do**
for $k \in \{1, ..., 5\}$ **do**
DailyTopics = LDA(T, k, method = "Gibbs", iterations = 1000)
$T_{DailyTerms}$ = (DailyTopics, 10)
Selecting top 10 terms from each topic

We apply the LDA in our dataset to detect the overall daily theme in the data, thus dividing our data on daily basis and perform the simple topic modeling with Gibbs sampling and 1000 iterations in order to identify the major events happened that day. The results of the topic modeling are shown below in the tables representing topics from each day. The key observation here is that from day 1 and day 2, the topic modeling does not provide us any word related to shooting or about the shooting event. However, day 3 onward does provide us with the topics such as "*shooting*", "*gun*", "*victims*", "*killed*", "*prayers*" etc. which are clearly related to the shooting event, day wise result of topic modeling shown in Table 1 for the first three date. Topic modeling does not help us in finding out the related keywords and topics to the shooting event from the day one. The emergency first responders need the information right from the very beginning of the event/day in order to rescue or provide appropriate help to the public. Thus, Topic modeling alone is not sufficient for the first responders, we need to include the sentiment analysis and change point detection algorithm in order to provide the timely and correct information about the event.

4.2. Sentiment and emotion analysis

The emotion extraction and analysis is the core module of this work which eventually leads towards the event detection thus, it is extremely important to get the high quality and large quantity emotion lexicon for better, reliable, and precise emotion detection. Emotion lexicons have obtained much less attention than general sentiment lexicon (positive and negative polarity) so there exist a lack in deep emotion lexicons and related dataset. There are a lot of information that can be captured by details emotional analysis similar to this work but at the same time we need to build such dataset to train and learn new models with higher accuracy. We have adopted an unsupervised approach (Lexicon/dictionary based) for sentiment and emotion classification thus we did not report the training, validation and testing set data along with their respective classification accuracy among other evaluation metrics on our data. The Lexicon (NRC Emotion Lexicon in this case) is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing and the function present in "syuzhet" package assess the sentiment and emotion of each word present in the sentence/tweet. The lexicon contains 14,182 unigrams (word-sentiment) and 25,000 senses (word-emotion). The NRC lexicon categorizes words in a binary fashion ("yes"/"no") into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

There are various works [28–31] using the word-embedding for lexicon expansion and enhancement in their model and using pre-trained models and different word-embedding [33,34]. In this section we propose to use a retrofitting algorithm [35] to enrich the lexicon knowledge and their semantic relationships using word embedding vectors in order to increase the size of our existing emotion lexicon, which can significantly improve the quality of our sentiment and emotion analysis module. Some of the popular resources for word embedding are word2Vec, Glove, ELMo etc. which contains the learned word vector representation from billions of word corpus such as Wikipedia, Google news.

Motivation for Expanding Emotion Lexicon: Rich, expanded and accurate lexicon contains the potential to improve the quality of the emotion detection. Since there can be application specific scenarios which prefer one particular kind of emotion lexicon (sad, angry, happy, disgust) over others such as (positive, negative, neutral) or (joy, love, suffering, surprise). Applications focused on detecting specific kind of emotions can utilize the expansion of the lexicon using various word embedding models that contains billions of words for better and reliable emotion detection.

4.2.1. Enriching word vectors using emotion lexicon

In this study we have used NRC Emotion lexicon (EmoLex) [38]. The data collection and evaluation provided in this paper [38] shows that small to moderate-sized emotion lexicon can be created effectively by crowdsourcing and Mechanical Turk under a certain budget but we need effective automation process to increase the size of such emotion lexicon. EmoLex contain 14,182 unigrams for sentiments (positive, negative) ad 25,000 senses that evoked by emotional terms (eight different emotion) selected from various sources including a known dataset WordNet [39]. Word embedding are the word vectors representations produced by popular algorithms in an unsupervised manner which made them such a powerful tool to be utilized for various task. Word representations in any vector space are learned from distributional information of words present in the large corpora and very helpful in providing the semantic information. Although there are word vectors such as Glove and Word2Vec but specific applications can improve their model's accuracy by combining these word representation with other lexicons such as WordNet and EmoLex which contains more semantically supported information. This work [35] proposes a retrofitting method to reconstruct the word vector representation space that utilize the relational information of other lexicons by supporting related, linked words with similar semantics and representation.

4.2.2. Retrofitting with semantic emotion lexicons

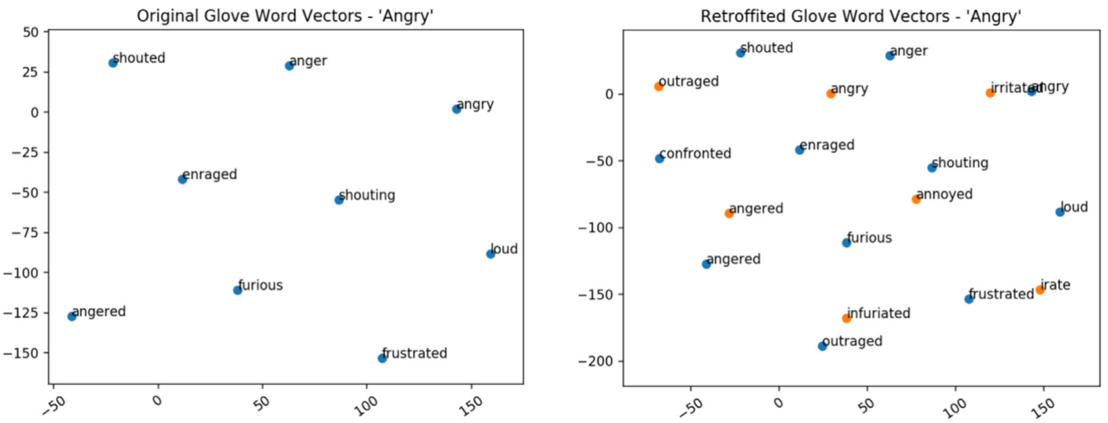
Let $V = w_1, \ldots, w_n$ be a vocabulary (set of word types) and an ontology Ω which encodes the semantic relations among words of vocabulary V. An undirected graph $(w_i, w_j) \in E \subseteq V * V$ represented a vertex as word type and edge as the semantic relationship of interest. Euclidean distance is used here as the distance measure between any pair of vectors. The goal is to learn the matrix $Q = (q_1, \ldots, q_n)$ from a distance measure where the columns are close to their counterpart $\hat{Q}$ in the adjacent vertices of Q in Ω.

The purpose is to minimize the distance and keep inferred word vector near to the observed value $\hat{q}_i$ and its neighbor q_j in the refine word vector space representation $\forall j$ such that $(i, j) \in E$. The objective function is as

$$\Psi(Q) = \sum_{i=1}^{n} \left[\alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j)\in E} \beta_{ij} \|q_i - q_j\|^2 \right] \tag{1}$$

Table 1
Top five topics and terms for Day 1, Day 2 and Day 3 (October 1st, 2nd and 3rd) using LDA.

Day 1					Day 2					Day 3				
Topics 1	Topics 2	Topics 3	Topics 4	Topics 5	Topics 1	Topics 2	Topics 3	Topics 4	Topics 5	Topics 1	Topics 2	Topics 3	Topics 4	Topics 5
great	free	see	vegas	the	see	vegas	free	more	the	victims	shooter	shooting	gun	now
oct	club	more	las	prison	prison	las	club	just	great	vegas	attack	the	just	congress
new	strip	just	now	years	good	now	strip	oct	new	las	massacre	will	need	you re
need	pick	good	night	home	years	home	pick	want	get	prayers	people	mass	like	weapons
real	call	want	today	latest	live	this	call	today	night	families	hotel	news	one	ban
this	credit	memories	were	thanks	tomorrowucufef	need	limo	october	entrepreneur	help	what	latest	people	assault
humidity	limo	entrepreneur	video	here	memories	real	have	cant	time	thoughts	room	killed	dont	demand
temperature	have	tomorrowucufef	saturday	lasvegassun	your	you	credit	help	one	our	paddock	trump	how	outraged
kmh	live	cant	you	morning	here	latest	code	humidity	its	tragedy	many	pray	time	please
team	code	time	will	day	morning	were	recommend	play	lasvegassun	love	shooters	violence	this	blood



(a) Glove vector representation for word Angry (b) retrofit vector representation for word Angry

Fig. 4. Word vector representation.

Here α and β controls the relative strengths of word vector associations. This process train the word vectors first independently without the semantic lexicon information then retrofit them with semantic lexicon information. Ψ is a convex in Q here and solved by a system of linear equations. Initially vectors in Q and $\hat{Q}$ are equal followed by taking the first derivative of Ψ w.r.t one q_i vector and equating it to zero to get the following online update

$$q_i = \frac{\sum_{j:(i,j)\in E} \beta_{ij} q_j + \alpha_i \hat{q}_i}{\sum_{j:(i,j)\in E} \beta_{ij} + \alpha_i} \tag{2}$$

This retrofitting approach can be applied to word vector representations obtained from any model as the updates in the above equation are not affected by the primary vector training model objective. The ontology Ω has the information about the semantic relations in the lexicon and it plays a crucial role towards modifying the learning objective of the vector training model as a prior. It is defined as

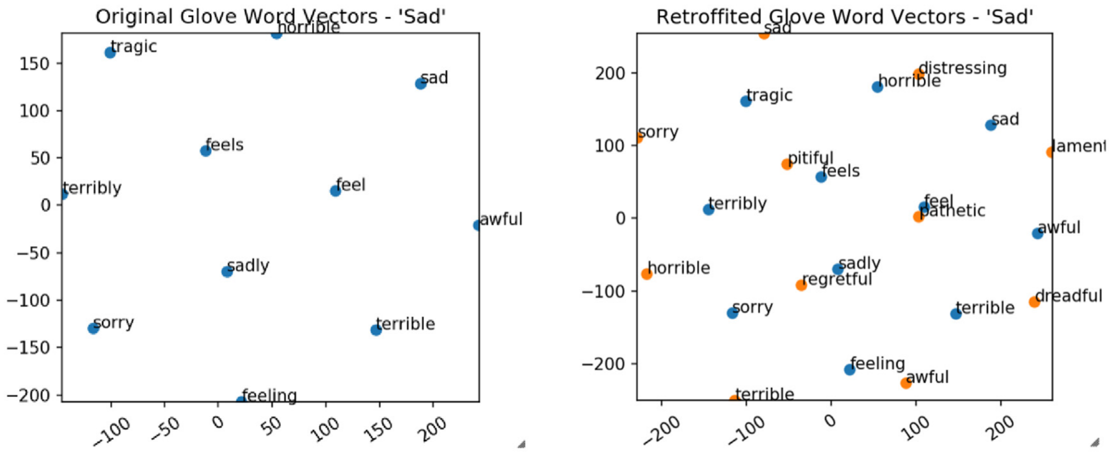
$$p(Q) \propto \exp -\gamma \sum_{i=1}^{n} \sum_{j:(i,j)\in E} \beta_{ij} \|q_i - q_j\|^2 \tag{3}$$

Here, strength of the prior is controlled by the hyperparameter γ. This prior encourages the connected words in the lexicon to have the close vector representation. Learning through estimation of maximum a posteriori can incorporate this prior by using the sum of log-likelihood gradients and log-priors (from above equations) w.r.t. Q provided by extant vector learning model for learning and updating after every k words.

Algorithm 2 Enriching The Emotion Lexicon

procedure EMOLEXEXPANSION(
 Inputs: Vocabulary V in *EmoLex*, Ontology Ω in V, Word Embedding $\hat{Q}$ learned from Glove
 Output: Expanded Lexicon (*EmoLex*+))
 Emotion Lexicon Expansion
 for k words **do**
  Train Word Vector $\hat{Q}$ independently through MAP using Eq. (3),
  Retrofit Vector Representation Q using Euclidean Distance measure $\Psi(Q)$ using Eq. (1),
  Update using $q_i \in Q$ using Eq. (2),
  Append to Q
 Extract Similar words (nearby) from Q in $EmoLex_{new}$
 Add in the existing lexicon i.e $EmoLex+ = EmoLex + EmoLex_{new}$

This method would add more words and enrich the lexicon with more semantically related words explained in Algorithm 2 resulting in increased emotion lexicon. This increase in lexicon size would enable the better emotion detection and improve the model performance.



(a) Glove vector representation for word Sad (b) retrofit vector representation for word Sad

Fig. 5. Word vector representation.

4.2.3. Expanding emotion lexicon with semantically similar words

Adding more semantically and sentimentally related words from other resources would enrich the lexicon without manual annotation and expansion in the size of the existing lexicon would improve the quality of the emotion extraction. One example can be seen in Fig. 4 where Fig. "4(a) shows the tSNE plot of the nearby original word vectors of word "Angry" and Fig. 4(b) shows the retrofitted near by word vectors of word "Angry".

Another example can be seen in Fig. 5 where Fig. 5(a) shows the tSNE plot of the nearby original word vectors of word "Sad" and Fig. 5(b) shows the retrofitted near by word vectors of word "Sad". These plots shown in Figs. 4(a) and 5(a) represents the example of only Glove vector representation for word "Angry" and "Sad" respectively whereas, Figs. 4(b) and 5(b) shows the updated lexicon that are mapped and semantically linked with the existing Glove word vectors. The count of words in Figs. 4(b) and 5(b) has been increased significantly (shown visually for word "angry" and "sad"). It also corroborates that the size of the lexicon with this retrofitted method increases resulting an enriched lexicon with dense word vector representation which can be utilized for further analysis effectively. More semantically related words can be added and manual annotation can be reduced for specific lexicons to expand the lexicon.

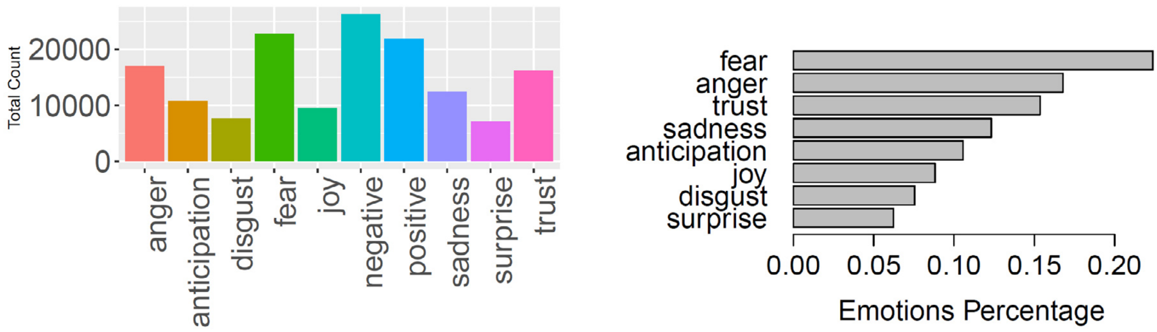
It is possible for some words to be present more than once such as "angered" and "sad" in Figs. 4(b) and 5(b) respectively. Since the data is coming from two different sources and context so there are high chances that some words might repeat and make the final outcome redundant. We can avoid it by only considering the non duplicate words in our final step of the algorithm.

4.3. Sentiment and emotion analysis over time

The sentiments and emotions are detected and classified with the help of R package "*syuzhet*" that includes "*NRC*" library developed by Mohammad, Saif M. and Turney, Peter D. in the NLP group at Stanford. This library calls the NRC Word-Emotion Association Lexicon to compute the presence of eight different type of emotions ("*anger*", "*anticipation*", "*disgust*", "*fear*", "*joy*", "*sadness*", "*surprise*", "*trust*") and two sentiments ("*positive*" and "*negative*") in the input text file. This Lexicon was created by following a manual method by crowdsourcing on Mechanical Turk. The terms present in this lexicon are chosen carefully that includes some frequent nouns, verbs, adjectives, adverbs, words from General Inquirer and some words from WordNet Affect Lexicon for fair comparison. This Lexicon contains words from general domain and labeled by Amazon Mechanical Turk to build this NRC word-association lexicon. It is based on unigrams and association scores works in binary mode (yes/no). It finally contains 14,182 unigrams (words) and almost 25,000 senses (eight different emotions) [38,40].

We used a lexicon based approach (unsupervised method) to classify our data into eight different emotions. In lexicon-based classification, data instances are assigned sentiment/emotion labels by comparing the words that appear in the lexicon with the input data and provide the label as binary form (yes/no i.e. belong to particular sentiment/emotion or not). The Fig. 6 shows the total emotions description present in the data. Fig. 6(a) shows the bar graph with different distribution of the emotions exist in the data which provides the insight about the emotion of the public about the event. Fig. 6(b) provides the amount of emotion in terms of percentage data. It showcases that emotion "*fear*" is the maximum followed by "*anger*" in the data which can be understood as the shooting was a terrible event.

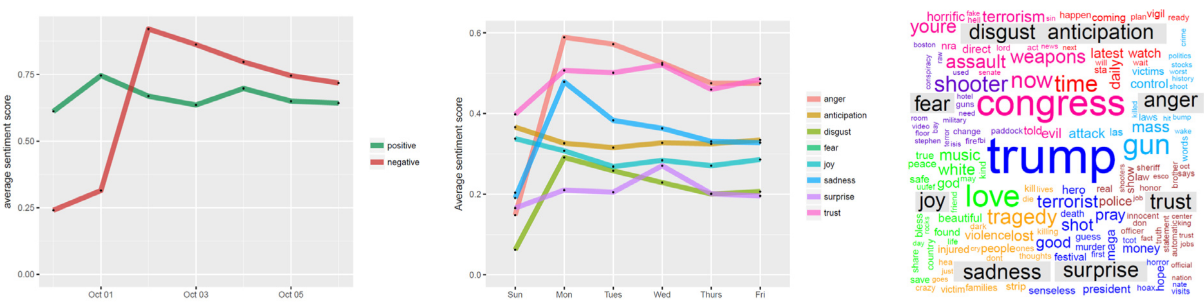
Fig. 7(a) shows the total sentiment polarity with time in the data which is helpful in determining the overall sentiments' polarity changing with time. The negative sentiment is increasing just after the horrible shooting occurred and as the



(a) Emotions distribution in Data

(b) Emotions percentage in data

Fig. 6. Total emotions in data.



(a) Sentiment polarity in the data with time

(b) Different Emotions Polarity in the Data

(c) Comparison Word Cloud of the Data

Fig. 7. Sentiment analysis visualization.

news spread more users were feeling bad and condemning the act which is the result of negative sentiments in users tweets. Though it is clear that negative sentiments are greater than the positive but it is still not helpful in forming any conclusive statements about the exact public sentiments and opinions. The negative sentiments could indicate various things including anger and fear but detailed emotions would be much more helpful. Furthermore, the plot in Fig. 7(b) shows the different emotion changing with the time. This time plot gives the crucial understanding of different emotions type and their respective intensity with time. The emotions' intensity can increase or decrease overtime but when this is happening could be significantly important to the local officials and emergency responders at that time. The changing nature of these emotions can be linked to the services provided that time and news available during the event. These links can worked as the constructive feedback to the governments and other responsible agencies to work in more effective manners. These different emotions can also be interpreted by the officials that people liked the service during and post event or there could be more improvements towards it.

The comparison cloud is generated to have a deeper understanding seeing the most frequent words that fall into each emotions type. The Fig. 7(c) shows the comparison cloud with all the emotion types. Also, different emotions captures different words and many words fall into two or more emotions resulting in overlapping tweets with other emotions. There are also some cases in which repetition of some words exist in different emotions but they might have used in different context. For example word "*Shooting*" used in Angry and Fear emotion can be seen in Table 2.

We also categorized the tweets based on their emotions to understand the overlapping of tweets in different emotions and the different distributions of tweets in each emotion category resulting as fear having 8000 tweets, angry having 12,000 tweets, anticipation having 8000 tweets, disgust having 6000 tweets, joy having 6000 tweets, sad having 9000 tweets, surprise having 5000 tweets and trust having 11,000 tweets approximately. It seems there are a lot of overlapping among some categories but we will address this in our future work.

4.4. Change point detection in various emotions

As it can be seen very clearly from Fig. 7(b) that emotions are changing with time and we are extremely interested in capturing those changes. The changes present in the sentiments can help in detecting the significant change points using the Bayesian change point detection (BCPD) algorithm introduced by Barry and Hartigan in 1993 [41] that uses the

Table 2
Tweets.

No.	Angry tweets	No.	Fear tweets
1	UPDATE Statement from VP fired for shameful comments about shooting victims	1	Churches help survivors prepare for longterm needs of more than people traumatized by shooting
2	Death toll in makes it the deadliest shooting in US history Other major shootings Orlando Virginia Tech	2	killed amp injured in shooting Thoughts n prayers arent enoughWorld MUST stand up against terror
3	Horrified to hear about the shooting in My thoughts are with the victims and their families Praying for everyone	3	BreakingWoman told group of concert goers youre all going to die tonight minutes before shootinghtt
4	I get sick watching the videos of the shooting in horrible disgusting amp heartbreaking	4	In light of the shooting here are ways to stay safe on amp off campus at Captioned
5	Looks like somebody is shooting at the crowds with a machine gun randomly	5	We need your help Anyone with photos or videos from shooting please contact our partners CALLFBI

product partition model for the normal errors change point problem using Markov Chain Monte Carlo. We used the R package for the BCPD [42] to find various change points in the input sequence. The change detection for sentiment can be represented as the univariate Change Point Detection problem in a Time Series Data.

Let $x(t)$ belongs to input data X (a single dimensional data generated at time t). To formulate the problem of change point detection in a time series data by assuming that $y(t)$ is a single dimensional time series instance of data generated at the time t. The algorithm uses a unknown partition $\rho = (U_1, U_2, \dots, U_n)$ into continuous blocks, where $U_i = 1$ indicates a change point at position $i+1$ and the transition between blocks are known as "*change points*". It initializes U_i to 0 for the positions $i < n$, with $U_n \equiv 1$. A value of U_i is computed from the conditional distribution of U_i, from the data and the current partition in every step of the Markov chain at each position i. Also, p as the probability of a change point at position i, independently for all i [42].

The change point detection algorithm assumes that the input observations X_i are independent $N(\mu_i, \sigma^2)$, where this notation specifies that μ_i are equal for all i within a block. The prior distribution of μ_{ij}, which means that the mean of the block beginning at position $i+1$ and ending at position j, is chosen as $N(\mu_0, \sigma_0^2/(j-i))$. One key observation in this process is that larger deviations from μ_0 are expected when there are shorter blocks, but weak signals can also be detected when adequate amount of data are available. w is defined as $\frac{\mu^2}{\mu^2+\mu_0^2}$ for convenience and $\mu_i \equiv \mu_{ij}$ for all i in block ij to avoid confusion. The transition probability, p, for the conditional probability of a change point at the position $i+1$, may be obtained from the simplified ratio presented by Barry and Hartigan

$$\frac{p_i}{1-p_i} = \frac{P(U_i = 1|X, U_j, j \neq i)}{P(U_i = 0|X, U_j, j \neq i)} = \frac{\left[\int_0^{\gamma} p^b (1-p)^{n-b-1} dp\right]\left[\int_0^{\lambda} \frac{w^{b/2}}{(W_1+B_1 w)^{(n-1)/2}} dw\right]}{\left[\int_0^{\gamma} p^{b-1} (1-p)^{n-b} dp\right]\left[\int_0^{\lambda} \frac{w^{(b-1)/2}}{(W_0+B_0 w)^{(n-1)/2}} dw\right]} \tag{4}$$

Here, b denotes the number of blocks obtained if $U_i = 0$, conditional on U_j, for $i \neq j$, W_0, B_0, W_1 and B_1 are the within and between block sums of squares obtained when $U_i = 0$ and $U_i = 1$ respectively, and X is the data. The tuning parameters γ and λ may take values in [0, 1]. The calculation of $f(U_i = 1|X, U_j, j \neq i)$ is done and sample of U_i taken at random. Once pass is completed of the data X the calculation of posterior mean μ_i at a given current position p_i is calculated using the formula

$$\hat{\mu_r} = (1-W)\hat{X}_{ij} + x\mu_0 \tag{5}$$

The tuning parameters μ and λ may take values in the range [0 : 1] and takes care the variation in the data thus, selected such that this method is stable and effective. After each iteration, the posterior means are updated conditional on the current partition. At last, after M passes the average of the M estimates of p_i is done by approximating the posterior mean p [41].

This recent study [26] has also shown that change point detection is useful in identifying potential sub-events that causes the changes in sentiments of general public over Twitter using the Delhi Election 2015 tweets data. The research has used the average sentiment score (positive and negative) and applied the change point detection algorithm to see the significant changes but in this research we are more focused in identifying the changes in every emotion.

The Algorithm 3 describes the complete procedure step by step. It takes the input as the raw tweets (T) with timestamp but these timestamp are not always equidistant here and returns the locations of the change points (L) as output. Once we get the locations of the change points, the tweets are extracted from those change points location. Summarize the tweets with some more surrounding tweets to get the clear understanding of the event that caused the change in the emotion. These events along with the emotional context can be send to the EMS for better situation awareness in the crisis.

Algorithm 3 Emotional Change Detection Algorithm

procedure ChangingEmotion(
  Inputs: Raw Tweets ($T = (t_1, t_2, ...t_n)$),
  Output: Change Locations (L))
 Data Processing
 for All tweets $t_i \in T$ **do**
  e_i = *GetEmotions(t_i)*
 for All emotions $e_i \in E$ **do**
  BCP(E)
  λ and γ(tuning parameters) as 0.2
  Posterior Mean μ_i position p_i
  Transition Probability p = $\frac{p_i}{1-p_i}$ using Eq. (4)
  if p $\geq$ *threshold* **then**
   Change in emotion = True
   Location (L) = Extract Location of changes(L_j)
   Return Location L
 Change Detected Tweets(T_{new}) = Extract tweets from those location$L$
 Summarize Tweets to get the event $EVENT_{summary}$

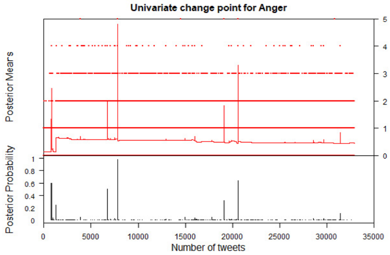
5. Results and analysis

In our research each emotion has a unique role in the event so by averaging the score of respective emotion, we would be loosing some important details about the emotions thus keeping all the emotion score we apply the BCPD on each emotion. Results shown in Fig. 8 include input as a sequence of emotional score and output as a values of posterior means and posterior probability for each emotion. The eight plots shown in Fig. 8 and each plot provides two plots that summarize the analysis of the Bayesian change point detection. The first (upper) part of the plot shows the Posterior mean that displays the data along with the posterior mean of each location in the number of tweets. The second (lower) plot displays Posterior probability of a Change which shows the proportion of iterations that results in a change point at each location in the number of tweets.

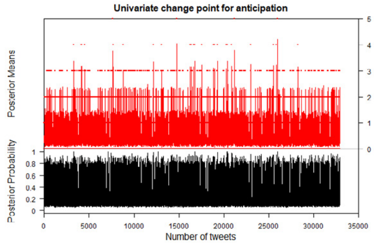
The spikes in the Posterior Probability part of the plot which are above certain threshold (say 0.4) can be considered as strong change points detected by the algorithm 3 and need appropriate attention in respective emotion category. Those locations of the change points and respective tweets along with few tweets above and below it are extracted from the original data to get the fair idea of which tweets are causing the change in emotions and can be summarized and sent to Decision makers along with the time stamp. Furthermore, tweets extracted from the location can be saved and summarized to address those issue in future emergency situations as well in order to handle them in more sophisticated way. We have controlled the threshold value using the optional parameters of Bayesian change point detection algorithm γ and λ by keeping it as 0.4, which is generally kept as 0.2 for better results but it can be varied dynamically depending on the granularity and how dense or sparse the change points are.

In Fig. 9(a), the change point detection for anger is expressed in terms of events that might be causing the changes with the emotion. In this process we manually checked from the ground truth information, the tweets at the change points detected from the algorithm representing the respective sentiments towards the event and the increase and decrease in the sentiments are captured nicely with the change point detection algorithm. In anger the first change point detected when many people posted about the O J Simpson (armed robber and kidnapper) got out from jail on parole after 9 years thus many people expressed their anger and fear towards the event. The second and third change points are mainly capturing the events related to dead and injures people because of shooting, the very high spike shows that people were very angry at that point because of the news of killing and injures victims all over the place. The last two change points in anger shows the events related to the gun violence and tragedy affecting people's life and no strong action taken from the president Trump.

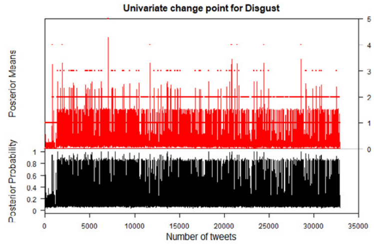
Another emotion fear is shown in Fig. 9(b) with events that could possibly cause the change in the emotion fear. The first few change detection are about the same as anger one, "*O J Simpson free from prison*", there were news and many people re-tweeted about the news and fearful about the result of his coming out of jail on parole. The next change detection in the same emotion is about the people expressing their feeling about the killings and other victims and their families, praying for each other, getting the news about the number of people killed and injured in the shooting. The ERS personnel can take advantage of this smart service system by looking at activities, experience and summary of this prior emergency event.



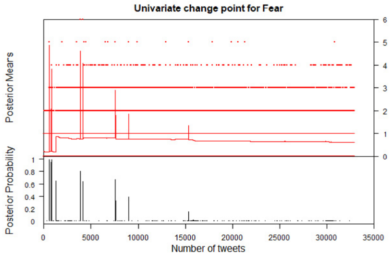
(a) Anger



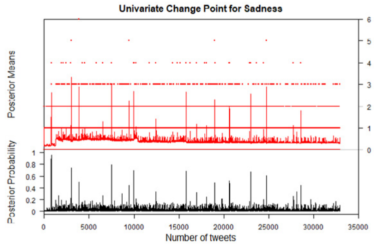
(b) Anticipation



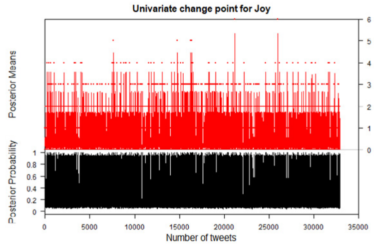
(c) Disgust



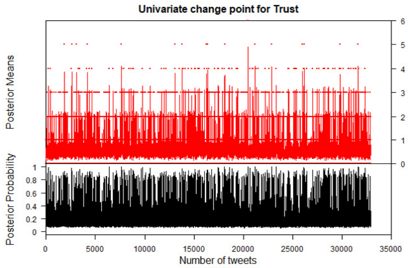
(d) Fear



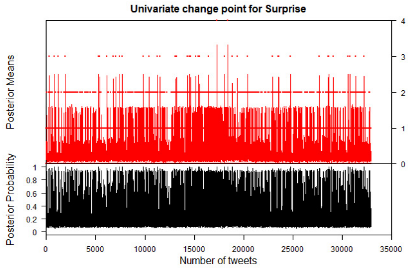
(e) Sadness



(f) Joy

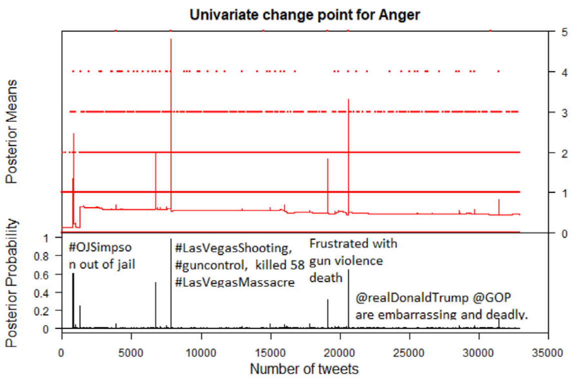


(g) Trust

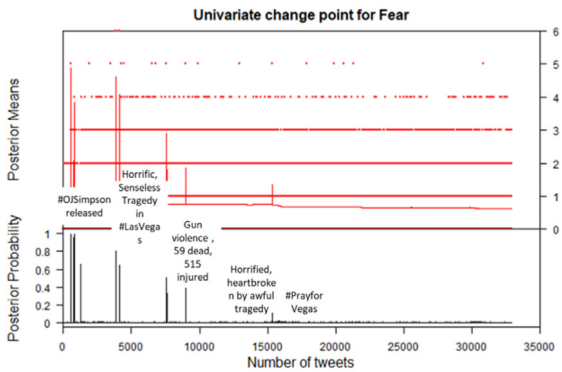


(h) Surprise

Fig. 8. Change point detection for different emotions.



(a) Events indicating change in emotion anger



(b) Events indicating change in emotion fear

Fig. 9. Events from emotional change detection.

5.1. Case study on a different data to demonstrate the generalization of our process

We have also performed the similar analysis on a different dataset and we have shown all the steps with their respective results in this section. We have followed the same steps as we did in our study to demonstrate the generalization of this approach.

Data Collection: This dataset is about another shooting event happened in Florida. The data collection has almost 280,000 tweets from February 13th, 2018 to February 19th, 2018 about the event that happened on February 14th, 2018 at the Florida Parkland high school, where at least 17 people were killed during the shooting. Fig. 10 shows the tweet counts per day and the tweet distribution in the new data collection.

Data Preprocessing, Cleaning and Visualization: We have performed the data cleaning, pre-processing and visualization steps explained in Sections 3.2 and 3.3 to generate these results shown in Fig. 11.

Emotion Analysis: Emotion analysis is the center of this study and we have performed the unsupervised lexicon based approach to detect the sentiment and emotions present in the data. Fig. 11(a) shows the bar graph of total emotions with their distribution in the data , Fig. 11(b) depicts the emotion comparison cloud for visual insights about different emotions and their associated word. Fig. 11(c) shows the positive sentiment decreasing and negative sentiment increasing with time since the horrific shooting happened on February 14th, 2018. Fig. 11(d) shows the emotions over time showing the sharp increase in emotion anger and fear, and sharp decrease in emotion joy and surprise (similar to our previous emotion analysis results).

Emotional Change Point Detection: As seen in Fig. 11(d) that fear, anger and joy emotions are changing significantly with time hence we perform the emotion change point detection i.e. Algorithm 3 to capture these changes and Algorithm 2 to perform the lexicon expansion. The following results are obtained by changes in the sequence of certain emotion score for respective emotion and output has the values of posterior means and the posterior probability in the plot Fig. 12. The eight plots shown in Fig. 12 and each plot provides two plots that summarize the analysis of the Bayesian change point detection.

Event Detection with Emotional Change Points: In Fig. 13(a), the change point detection for anger is expressed in terms of events that might be causing the changes with the emotion. In this process we manually checked from the ground truth information, the tweets at the change points detected from the algorithm representing the respective sentiments towards the event and the increase and decrease in the sentiments are captured nicely with the change point detection algorithm i.e. Algorithm 3. In Fig. 13(a) based on the change point locations people have expressed their anger about shooting that happened in Parkland School, Florida and tweeting about Gun Control and reforms etc. In Fig. 13(b) emotional changes due to fear in the general public mainly focused on deadly shooting, assault rifle, gun violence and killing of 17 students. This tragic shooting was prime example of hatred and evil act tweeted by many people as well.

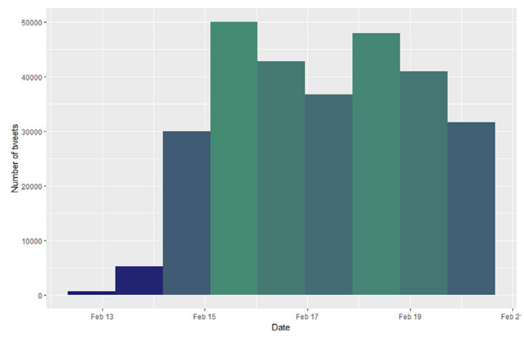
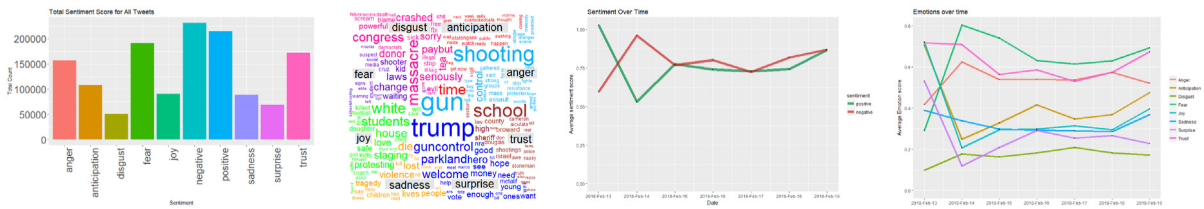
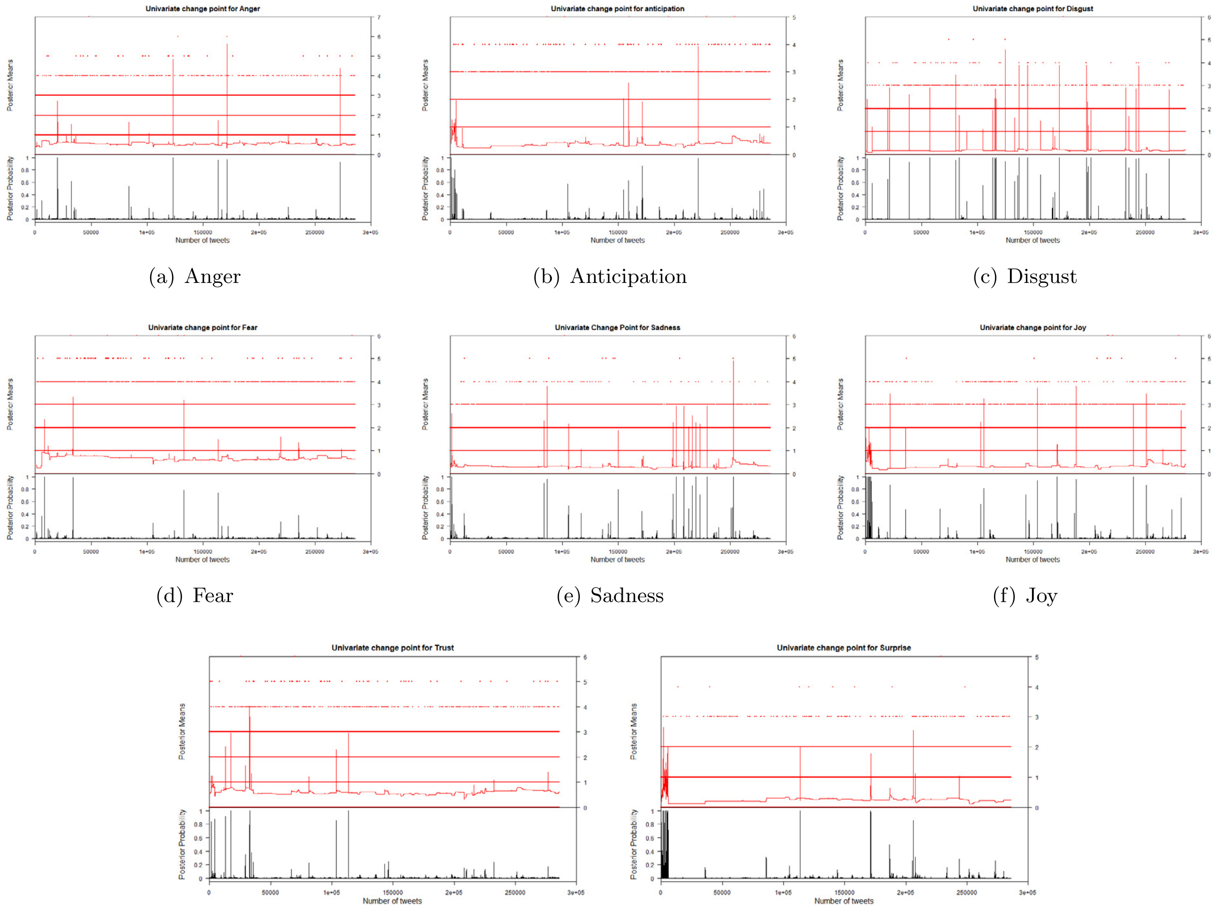


Fig. 10. Tweets count in data.



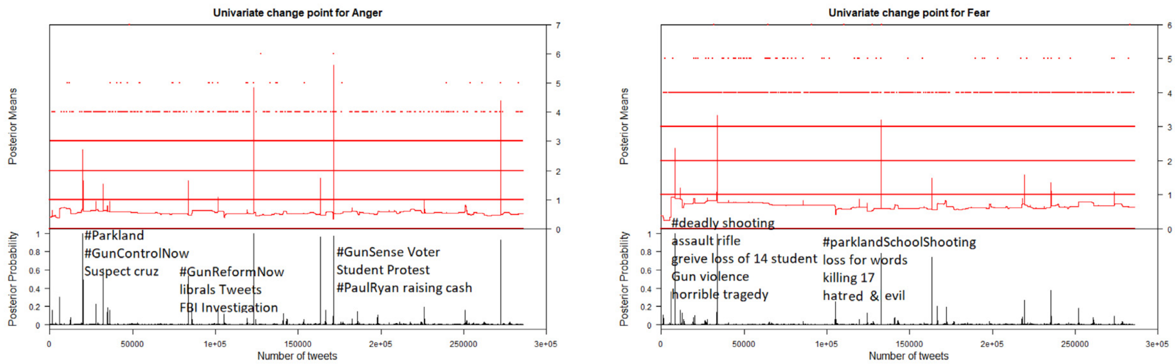
(a) Total Emotions Present in Data (b) Comparison Word-Cloud (c) Sentiment Over Time (d) Emotions Over Time

Fig. 11. Emotion analysis of the example data.



(a) Anger (b) Anticipation (c) Disgust (d) Fear (e) Sadness (f) Joy (g) Trust (h) Surprise

Fig. 12. Change point detection for different emotions.



(a) Events indicating change in emotion anger (b) Events indicating change in emotion fear

Fig. 13. Events from emotional change points.

Comparison between Las Vegas and Florida dataset analysis: As we have seen that the results obtained by our primary data (Las Vegas Shooting) and example data (Florida Shooting) analysis in this study are quite similar. These alike results and analysis provides the evidence that this research process can be applied to comparable situations and can be generalized. For example, Figs. 7(a) and 11(c) shows the positive sentiment decreasing and negative sentiment increasing with time on the day of shooting. Emotion anger, fear increasing and joy decreasing in Figs. 7(b) and 11(d), given the horrific shooting event. Events are also identified with the help of emotion change point detection in emotions anger and fear shown in Figs. 9 and 13.

6. Challenges and limitation

In this section we briefly summarize the underpinning challenges we faced.

6.1. Challenges using emotion detection

In this study we are not utilizing all the emotions that are detected through sentiment analysis and focusing on dominating emotions. The reason behind this, every event or topic generally have some dominating emotions involved, for example if the event is related to crisis, people tend to express the emotions such as sadness, anger, fear etc., more often than joy and surprise. If the event is political or social issue, we see all kinds of mixed opinion and emotions. Thus, the event type greatly influence the kind of emotions might be produced by the public. In this study, we are using the EmoLex lexicon to get the eight different emotions but there are many other lexicon present to compare which lexicon will be suitable for such events. The evaluation of this dictionary based method could not be compared with some machine learning based techniques mainly because of manual annotations used therein.

6.2. Challenges using change detection with different emotions

As we can see from Fig. 8 that change detection in the emotion Anger and emotion Fear is quite sparse in comparison to the other emotions while other emotions have very dense representation of change point. Based on the information gathered from Fig. 7(b), we can say that the emotion Anger and Fear have the similar kind of representation thus showing a significant change at specific location while other emotions have different representation, showing numerous change points at multiple locations. Also, emotion Joy, Disgust and Anticipation have a very dense representation of change points, making it difficult to figure out the actual changes. The data distribution of the emotions is generally sparse which makes the change point detection algorithm perform really bad for some emotions and good for the dominating emotions.

7. Conclusion and future work

In this study, a small dataset of Las Vegas Shooting happened in October 2017, is collected and visualized. Data is analyzed based on the number of tweets collected per day and the sentiments involved in the data. Sentiment polarity (positive and negative) is identified in the data but it is not good enough to understand the detailed emotions of the public during the crisis event. Furthermore, eight different types of emotions are identified in the dataset to get the deeper insights of the event that can be helpful for the local authorities and emergency responders to make the event handling more emotional aware. The emotions detection and tracking the changes of these emotions overtime will be more meaningful rather than tracking the event alone. Bayesian Change point detection algorithm is applied on the eight different emotions in order to identify the changes in the emotions. The locations of the change point detection are identified and the respective tweets are extracted to locate the cause of the change in emotions. Once the change point tweets and their respective and surrounded tweets are summarized to provide the context of the emotion change and the events caused by that changes. Emotion detection along with the context of the emotional change can be identified and understood by the authorities to get the better situation awareness and deal the crisis in enlightened way. We have also applied the emotion change point detection algorithm on a different dataset (Florida Shooting) to get the insights of the events caused by emotion change through this study and shown the generalization of this approach.

In this work we have used the NRC Emotion lexicon (EmoLex) based (unsupervised) approach but we would also like to experiment with the supervised machine learning based approach (with labeled data) and deep learning approaches to get the specific accuracy and robust model. We would like to compare this approach with other lexicon based sentiment analysis model for fair comparison, address the correlation among all the emotion and access the impact of it. We would also like to automate the event detection part of this process for better visualization and decision making for emergency response services.

Acknowledgment

This work is partially supported by the National Science Foundation (NSF), United States CNS grant 1640625.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Kwak, C. Lee, H. Park, S.B. Moon, What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010, 2010, pp. 591–600, http://dx.doi.org/10.1145/1772690.1772751.
[2] A. Crooks, A. Croitoru, A. Stefanidis, J. Radzikowski, #Earthquake: Twitter as a distributed sensor system, Trans. GIS 17 (1) (2013) 124–147, http://dx.doi.org/10.1111/j.1467-9671.2012.01359.x.
[3] D. Wang, M.T.A. Amin, S. Li, T.F. Abdelzaher, L.M. Kaplan, S. Gu, C. Pan, H. Liu, C.C. Aggarwal, R.K. Ganti, X. Wang, P. Mohapatra, B.K. Szymanski, H.K. Le, Using humans as sensors: an estimation-theoretic perspective, in: IPSN'14, Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (Part of CPS Week), April 15–17, 2014, Berlin, Germany, 2014, pp. 35–46.
[4] M. Kranz, L. Roalter, F. Michahelles, Things that twitter: social networks and the internet of things, in: What Can the Internet of Things Do for the Citizen (CIoT) Workshop At the Eighth International Conference on Pervasive Computing (Pervasive 2010), 2010, pp. 1–10.
[5] G. Sagl, B. Resch, B. Hawelka, E. Beinat, From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments, in: Proceedings of the GI-Forum, 2012, pp. 54–63.
[6] F. Atefeh, W. Khreich, A survey of techniques for event detection in twitter, Comput. Intell. 31 (1) (2015) 132–164.
[7] S. Wang, P. Giridhar, L. Kaplan, T. Abdelzaher, Unsupervised event tracking by integrating twitter and instagram, in: Proceedings of the 2nd International Workshop on Social Sensing, ACM, 2017, pp. 81–86.
[8] P. Giridhar, S. Wang, T. Abdelzaher, Event tracker using social networks: Demo abstract, in: Proceedings of the Second International Conference on Internet-of-Things Design and Implementation, ACM, 2017, pp. 285–286.
[9] P. Giridhar, T. Abdelzaher, J. George, L. Kaplan, On quality of event localization from social network feeds, in: Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on, IEEE, 2015, pp. 75–80.
[10] B. Pang, L. Lee, et al., Opinion mining and sentiment analysis, Found. Trends Inf. Retr. 2 (1–2) (2008) 1–135.
[11] D. Jurafsky, Speech and language processing: An introduction to natural language processing, Comput. Linguist. Speech Recognit. (2000).
[12] C.-H. Chang, M. Kayed, M.R. Girgis, K.F. Shaalan, A survey of web information extraction systems, IEEE Trans. Knowl. Data Eng. 18 (10) (2006) 1411–1428.
[13] C. Nord, Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-Oriented Text Analysis, No. 94, Rodopi, 2005.
[14] K.W. Church, R.L. Mercer, Introduction to the special issue on computational linguistics using large corpora, Comput. Linguist. 19 (1) (1993) 1–24.
[15] B. Pang, L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, p. 271.
[16] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, in: LREC, vol. 10, 2010, pp. 2200–2204.
[17] M. Etter, E. Colleoni, L. Illia, K. Meggiorin, A. D'Eugenio, Measuring organizational legitimacy in social media: Assessing citizens' judgments with sentiment analysis, Bus. Soc. 57 (1) (2018) 60–97.
[18] S. Sohangir, D. Wang, A. Pomeranets, T.M. Khoshgoftaar, Big Data: Deep learning for financial sentiment analysis, J. Big Data 5 (1) (2018) 3.
[19] V.K. Neppalli, C. Caragea, A. Squicciarini, A. Tapia, S. Stehle, Sentiment analysis during hurricane sandy in emergency response, Int. J. Disaster Risk Reduct. 21 (2017) 213–222.
[20] S. Goyal, Review paper on sentiment analysis of twitter data using text mining and hybrid classification approach, Int. J. Eng. Dev. Res. (2017).
[21] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113.
[22] D. Buscaldi, I. Hernandez-Farias, Sentiment analysis on microblogs for natural disasters management: a study on the 2014 genoa floodings, in: Proceedings of the 24th International Conference on World Wide Web, ACM, 2015, pp. 1185–1188.
[23] G. Beigi, X. Hu, R. Maciejewski, H. Liu, An overview of sentiment analysis in social media and its applications in disaster relief, in: Sentiment Analysis and Ontology Engineering, Springer, 2016, pp. 313–340.
[24] Y. Ouyang, B. Guo, J. Zhang, Z. Yu, X. Zhou, Sentistory: multi-grained sentiment analysis and event summarization with crowdsourced social media data, Pers. Ubiquitous Comput. 21 (1) (2017) 97–111.
[25] H. Shekhar, S. Setty, Disaster analysis through tweets, in: Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on, IEEE, 2015, pp. 1719–1723.
[26] R. Srivastava, M. Bhatia, Real-time unspecified major sub-events detection in the twitter data stream that Cause the change in the sentiment score of the targeted event, Int. J. Inf. Technol. Web Eng. 12 (4) (2017) 1–21.
[27] C. Truong, L. Oudre, N. Vayatis, A review of change point detection methods, 2018, arXiv preprint arXiv:1801.00718.
[28] H. Saif, M. Fernandez, L. Kastler, H. Alani, Sentiment lexicon adaptation with context and semantics for the social web, Semant. Web 8 (5) (2017) 643–665.
[29] G. Daval-Frerot, A. Bouchekif, A. Moreau, Epita at SemEval-2018 Task 1: Sentiment analysis using transfer learning approach, in: Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, pp. 151–155.
[30] X. Dong, G. Melo, A helping hand: Transfer learning for deep sentiment analysis, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, 2018, pp. 2524–2534.
[31] L.-C. Yu, J. Wang, K.R. Lai, X. Zhang, Refining word embeddings for sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 534–539.
[32] L.-C. Yu, J. Wang, K. Lai, X. Zhang, Refining word embeddings using intensity scores for sentiment analysis, IEEE/ACM Trans. Audio Speech Lang. Process. (2017) 1–1.
[33] G. Badaro, H. Jundi, H. Hajj, W. El-Hajj, EmoWordNet: Automatic Expansion of Emotion Lexicon Using English WordNet, in: Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, 2018, pp. 86–93.
[34] Y. Wang, Y. Zhang, B. Liu, Sentiment lexicon expansion based on neural PU Learning, double dictionary lookup, and polarity association, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 553–563.
[35] M. Faruqui, J. Dodge, S.K. Jauhar, C. Dyer, E. Hovy, N.A. Smith, Retrofitting word vectors to semantic lexicons, 2014, arXiv preprint arXiv:1411.4166.
[36] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (Jan) (2003) 993–1022.
[37] D.M. Blei, Probabilistic topic models, Commun. ACM 55 (4) (2012) 77–84.
[38] S.M. Mohammad, P.D. Turney, Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches To Analysis and Generation of Emotion in Text, Association for Computational Linguistics, 2010, pp. 26–34.
[39] G. Miller, WordNet: An Electronic Lexical Database, MIT Press, 1998.
[40] S.M. Mohammad, P.D. Turney, Crowdsourcing a word–emotion association lexicon, Comput. Intell. 29 (3) (2013) 436–465.
[41] D. Barry, J.A. Hartigan, A Bayesian analysis for change point problems, J. Amer. Statist. Assoc. 88 (421) (1993) 309–319.
[42] C. Erdman, J.W. Emerson, et al., Bcp: an R package for performing a Bayesian analysis of change point problems, J. Stat. Softw. (2007).