UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

JONATHAS GABRIEL DIPP HARB

Using a convolutional neural network to compare emotional reactions on Twitter to mass violent events

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Prof[a]. Dr[a]. Karin Becker

Porto Alegre
May 2019

CIP — CATALOGING-IN-PUBLICATION

Dipp Harb, Jonathas Gabriel

Using a convolutional neural network to compare emotional reactions on Twitter to mass violent events / Jonathas Gabriel Dipp Harb. – Porto Alegre: PPGC da UFRGS, 2019.

77 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2019. Advisor: Karin Becker.

1. Deep Learning. 2. Convolutional Neural Network. 3. Sentiment Analysis. 4. Twitter. 5. Mass Violent Events. I. Becker, Karin. II. Título.

"For me life is continuously being hungry. The meaning of life is not simply to exist, to survive, but to move ahead, to go up, to achieve, to conquer.."

— Arnold Schwarzenegger

ACKNOWLEDGEMENTS

I thank God first for giving me the opportunity to start and finish this research, giving me health and strength in all moments.

My gratitude to everyone around me (close family, friends and colleagues), being surrounded by such amazing people pushes me to work harder every day.

Special thanks to my advisor Prof. Karin Becker, for her constant support and motivation in keeping me moving ahead with this research, and for all the teachings in those three years.

I am grateful to the Federal University of Rio Grande do Sul and the Institute of Informatics, for providing excellent infrastructure in those three years, and all the years before during my graduation, which allowed me to overcome limits and to become wiser, persistent and resilient to conclude this research.

ABSTRACT

The number of mass violent attacks, particularly mass shooting and terrorism events, has increased in recent years. Understanding the emotional reaction of the population is very important to help them to cope with the constant sense of threat and fear effectively. In this paper, we apply deep learning techniques to classify emotions expressed by Twitter users, and develop a comparative analysis of emotional reactions to twelve mass violent events, detailed using demographics information extracted from the users profiles. To classify the emotions, we trained a Convolutional Neural Network combining sets of automatically filtered training seeds and pre-trained word embeddings. Our study compared four terrorism events and eight mass shooting incidents in terms of emotional shift and prevalent emotions; influence on emotions of gender, age, closeness to the event and number/type of victims; and terms used to express reactions. We observed similar patterns for both kinds of events, mainly in terms of prevalent emotions (anger, fear, and sadness, respectively) and influence of gender on emotions (e.g. fear for women, and anger for men). The proximity to the events is influential only in mass shooting events. Tweeters expressing fear and sadness tend to share words of empathy and support, while those expressing anger tend to use intense words of hate, intolerance and call for justice.
Keywords: Deep Learning. Convolutional Neural Network. Sentiment Analysis. Twitter. Mass Violent Events.

Usando uma rede neural convolucional para comparar reações emocionais no Twitter de eventos violentos em massa

RESUMO

O número de ataques violentos em massa, particularmente de tiroteio em massa e terrorismo, aumentou nos últimos anos. Compreender a reação emocional da população é muito importante para ajudá-los a lidar com o constante sentimento de ameaça e medo de forma eficaz. Neste trabalho, aplicamos técnicas de aprendizagem profunda para classificar emoções expressas por usuários do *Twitter* e desenvolvemos uma análise comparativa de reações emocionais referente à doze eventos violentos em massa, detalhados usando informações demográficas extraídas dos perfis de usuários. Para classificar as emoções, uma rede neural convolucional foi treinada, combinando conjuntos de sementes de treinamento automaticamente filtradas e *Word Embeddings* pré-treinados. Foram comparados quatro eventos de terrorismo e oito incidentes de tiro em massa em termos de mudança emocional e emoções prevalentes; influência nas emoções devido a gênero, idade, proximidade com o evento e número/tipo de vítimas; e também termos usados para expressar as reações. Foram observados padrões semelhantes para os dois tipos de eventos, principalmente em termos de emoções prevalentes (raiva, medo e tristeza, respectivamente) e influência do gênero nas emoções (por exemplo, medo para as mulheres e raiva para os homens). A proximidade com os eventos é influente apenas em eventos de tiro em massa. *Tweeters* expressando medo e tristeza tendem a compartilhar palavras de empatia e apoio, enquanto aqueles que expressam raiva tendem a usar palavras intensas de ódio, intolerância e clamor por justiça.

Palavras-chave: Aprendizagem Profunda. Rede Neural Convolucional. Análise de Sentimentos. Twitter. Eventos Violentos em Massa.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AMT	Amazon Mechanical Turk
CNN	Convolutional Neural Network
FNN	Feedforward Neural Network
GloVe	Global Vectors
HIT	Human Intelligence Task
MS	Mass Shooting
ME	Maximum Entropy
NB	Naïve Bayes
NLP	Natural Language Processing
POS	Part-of-Speech
RNN	Recursive/Recurrent Neural Network
SVM	Support Vector Machines
TR	Terrorism (TR)
UK	United Kingdom
US	United States of America

CONTENTS

1 INTRODUCTION

The world has witnessed a growing number of violent mass attacks in the recent years, particularly in the form of mass shooting incidents and terrorism events. Broadly speaking, mass shooting (or mass killing) refers to an incident involving multiple victims of gun violence (AGGRAWAL, 2005). Individuals who commit mass shootings may fall into any number of categories, including killers of coworkers, students, and random strangers. Revenge is a notable motivation, but others are possible, including the need for attention, religious or political beliefs, discrimination, and mental illness. Mass shooting episodes have increased over recent decades and have received substantial media coverage (LOWE; GALEA, 2015). In the United States of America (US), mass shootings have claimed more lives these days than at any other point in the past thirty-five years[1].

Terrorism events are also increasing, and have reached a point where they are threatening the safety of the populace on a global scale (ROSER; NAGDY; RITCHIE, 2018). The Global Terrorism Database[2] defines a terrorist attack as the "threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation". According to this definition, GDT reports almost 15,000 terrorism-related incidents in 2016, a number three times higher than in 2010. These events are increasingly taking place in Western countries, particularly the United Kingdom (UK) and US. Hoffman (2013) defines terrorism as "the deliberate creation and exploitation of fear through violence or the threat of violence in the pursuit of political change". He stresses that terrorism events are specifically designed to have far-reaching psychological effects beyond the immediate victim(s) or object of the terrorist attack. Crenshaw (1981) argues that terrorism is simply a method that is used in the service of a cause, and breaks down the causes of terrorism into three layers: situational factors (e.g. conditions for radicalization or that motivate feelings against the ’enemy’), strategic aims (e.g. advertisement of a cause, influence a public attitude) and individual motivations (e.g. personality traits).

Terrorism and mass shooting incidents are similar in many respects. They involve planning, aim to reach a group of people with a single strike (possibly randomly selected), produce mass casualties, and target a public place. According to Crenshaw (1981), the main difference between terrorism and mass shooting events lies in motive. Given that a terrorism event is the action of a rational actor, the goal is to send a message through the act itself and to communicate the threat of another attack (CRENSHAW, 1981). In a mass shooting, the perpetrator does not

[1]https://www.newyorker.com/news/daily-comment/searching-for-motives-in-mass-shootings
[2]https://www.start.umd.edu/gtd/

wish to make a statement. According to Aggrawal (2005), a mass shooter kills in order to gain a brief moment of control by controlling the fate of others. Given these similarities, throughout this work we shall refer to these two types of events as *mass violent events*.

The research to date provides evidence that mass violent events can have mental health consequences for direct victims and members of affected communities, such as post traumatic stress disorder, depression, and other psychological symptoms (LOWE; GALEA, 2015; HOFFMAN, 2013). It is well known that mass violent incidents exert a psychological toll on their direct victims and members of the communities in which they took place, but social media and 24 hour news coverage of these attacks and their aftermath reach far beyond the affected communities to the entire nation and beyond (LOWE; BLACHMAN-FORSHAY; KOENEN, 2015). They impact people in a complex emotional way as their goal is to outreach many people in a single strike, resulting in a feeling of fear and insecurity to a larger portion of the population (MAGUEN; PAPA; LITZ, 2008). It has come to a point where the capacity of emotionally coping with such events has become a public health issue (LOWE; BLACHMAN-FORSHAY; KOENEN, 2015). This constant sense of fear and insecurity also increases the prejudice and intolerance towards certain ethnic or religious groups (e.g. muslims), as people tend to relate them to the radicals that perpetrated the attack. Understanding the emotional reactions of the population is very important to design assistance programs that effectively help the population to deal with these issues (CREPEAU-HOBSON et al., 2012; MAGUEN; PAPA; LITZ, 2008; COHEN-LOUCK; BEN-DAVID, 2017).

The study of the emotional impact of community-wide trauma is a complex and costly task. Jones et al. (2016) stress that most challenges are related to the data collection process, including inability to control the respondents' pre-event behavior (since the data collection starts after the event), timing for collecting data, as some victims tend to forget the symptoms they experienced by the event, and difficulty of getting access to or assembling participation from traumatized community members. Researchers are circumventing these challenges by exploring data extracted from social media as an alternative (JONES et al., 2016). Twitter is a popular social media platform used for posting real-time discussions, thoughts, sentiments and opinions with regard to several topics. For many events, Twitter has become the primary source of news, updates, and awareness (LI et al., 2017).

Sentiment Analysis deals with the automatic extraction and interpretation of people's opinions, attitude, and emotions from documents (LIU, 2012). Sentiment analysis was deployed in terrorism-related tweets to study the information diffusion model (BURNAP et al., 2014; GARG; GARG; RANGA, 2017; SIMON et al., 2014) and emotional contagion (CHONG, 2016), identify terrorism-related sentiment (AZIZAN; AZIZ, 2017; MIRANI; SASI, 2016), and compare sentiments of Western and Eastern populations towards a terrorism organization (MANSOUR, 2018). A study reveals the adequacy of Twitter-derived data to understand the psychological effects of mass shooting events on local communities (JONES et al., 2016). Another study (WANG; VARGHESE; DONNELLY, 2016) investigated pro/against gun stances related to a school mass shooting. A difference on sentiment expression was identified by comparing reactions to a mass shooting event and to two natural disasters (VARGAS et al., 2016). All of these works are restricted to *polarity* analysis, i.e. sentiment classified merely as positive and negative.

Sentiment analysis has been extended to *emotions*, such as the Ekman's basic emotions model (EKMAN; FRIESEN, 1982) (e.g. joy, anger, fear). Emotions have been examined on Twitter for understanding different social phenomena, such as structure of social connections, urban mobility patterns, participation in racial equality activism movements, and Gender-based violence, among others (LERMAN et al., 2016; GALLEGOS et al., 2016; CHOUDHURY et al., 2016; ELSHERIEF; BELDING; NGUYEN, 2017). To the best of our knowledge, the only work that addresses emotions in mass violent events is (SINGH; CHOUDHARY, 2017), which analyzes the basic emotions contained in tweets about a terrorism event that occurred in Barcelona using a sentiment lexicon.

There are two basic approaches for sentiment analysis: sentiment lexicons and supervised learning. Most aforementioned works deploy a lexicon-based approach (JONES et al., 2016; SINGH; CHOUDHARY, 2017; MANSOUR, 2018), which tends to not perform well due to the low recall. Supervised learning, on the other hand, is dependent on an annotated corpus, preferably domain-specific (LIU, 2012; ZIMBRA et al., 2018). Related work in the mass violence context relied on the engineering of features, as well on training datasets that were either manually created (WANG; VARGHESE; DONNELLY, 2016; VARGAS et al., 2016) or automatically labeled using lexicons (MIRANI; SASI, 2016; AZIZAN; AZIZ, 2017; GARG; GARG; RANGA, 2017). Deep learning refers to learning procedures that make use of computing power and large sets of data. It enables learning higher level abstractions from words and their context from huge volumes of data, not limited to annotated instances (e.g. pre-trained word embeddings) (ZHANG; WANG; LIU, 2018; COLLOBERT et al., 2011).

In this work, we study the emotional reaction of Twitter users to mass violent events. To this purpose, we apply deep learning techniques to develop an emotion classifier targeted at these types of event, and analyze the emotional reactions based on the demographics of the tweeters, particularly gender, age and location. We also investigate differences in how

Table 1.1: Overall Information on the Terrorism Events

Event Name	Location	Date	Type	Victims
#prayformanchester[a]	Manchester-UK	22 May 2017	Concert	22 deaths / 60 injuries
#londonbridge[b]	London-UK	3 June 2017	Others (public place)	8 deaths / 48 injuries
NY-october[c]	New York City-US	31 October 2017	Others (public place)	8 deaths / 11 injuries
NY-december[d]	New York City-US	11 December 2017	Subway	0 deaths / 4 injuries

[a]https://en.wikipedia.org/wiki/Manchester_Arena_bombing
[b]https://en.wikipedia.org/wiki/2017_London_Bridge_attack
[c]https://en.wikipedia.org/wiki/2017_New_York_City_truck_attack
[d]https://en.wikipedia.org/wiki/2017_New_York_City_attempted$_{b}ombing$

emotions are expressed depending on the event, and include a comparative analysis of emotional reactions to terrorism events and mass shooting incidents. The emotion classification model was developed by training a Convolutional Neural Network (CNN) (KIM, 2014) over 12 (twelve) datasets of tweets related to terrorism and mass shooting events, from which training seeds were automatically labeled. Preliminary results of this research were reported in (HARB; BECKER, 2018; HARB; BECKER, 2019).

Our analysis aims to answer the following research questions:

- Q1: Is there an emotion shift due to mass violent events?
- Q2: Do different mass violent events evoke the same emotional reaction?
- Q3: Does the proximity to the event influence the emotional reaction?
- Q4: Does the scale of casualties have an impact on the emotional reaction?
- Q5: Are there differences in the expression of emotions depending on the event?

Our study encompassed four (4) terrorism events and eight (8) mass shooting incidents. The terrorism events, described in Table 1.1, occurred in two distinct countries (UK and US), and two of them took place in the same city (New York City). They involved distinct targets and methods, and produced casualties of different scales. The mass shooting events are detailed in Table 1.2. They all took place in the US, and varied in target (e.g. school, church) and scope of casualties. All of these events drew a lot of attention of world news outlets and social media, and were largely publicized. Due to these properties, they are representative events and suitable for our comparative studies.

Our main contributions are:

- the development of a classifier for emotion identification in the context of mass violent events, by training a CNN with an automatically selected set of tweets as seeds, due to the lack of labeled datasets;

Table 1.2: Overall Information on the Mass Shooting Events

Event Name	Location	Date	Type	Victims
lasvegas[a]	Las Vegas - Nevada	01 Oct 2017	Concert	58 deaths / 851 injuries
marshallcounty[b]	Marshall County - Kentucky	23 Jan 2018	School	2 deaths / 18 injuries
maryland[c]	Annapolis - Maryland	28 Jun 2018	Newspaper Office	5 deaths / 2 injuries
orlando[d]	Orlando - Florida	12 Jun 2016	Nightclub	49 deaths / 53 injuries
ranchotehama[e]	Tehama County - California	14 Nov 2017	School and others	5 deaths / 18 injuries
santafe[f]	Santa Fe - Texas	18 May 2018	School	10 deaths / 13 injuries
stoneman[g]	Parkland - Florida	14 Feb 2018	School	17 deaths / 17 injuries
sutherland[h]	Sutherland Springs - Texas	5 Nov 2017	Church	26 deaths / 20 injuries

[a] https://en.wikipedia.org/wiki/2017_Las_Vegas_shooting
[b] https://en.wikipedia.org/wiki/Marshall_County_High_School_shooting
[c] https://en.wikipedia.org/wiki/Capital_Gazette_shooting
[d] https://en.wikipedia.org/wiki/Orlando_nightclub_shooting
[e] https://en.wikipedia.org/wiki/Rancho_Tehama_Reserve_shootings
[f] https://en.wikipedia.org/wiki/Santa_Fe_High_School_shooting
[g] https://en.wikipedia.org/wiki/Stoneman_Douglas_High_School_shooting
[h] https://en.wikipedia.org/wiki/Sutherland_Springs_church_shooting

- an analysis of the emotional reaction to mass violent events that encompasses the demographic information of Twitter users, extracted from their profile, as well as how emotions were expressed;
- a comparison of the emotional reactions related to each type of event;
- twelve (12) tweet datasets for this domain, and two (2) manually annotated Gold Standards, one for each type of event (i.e. terrorism and mass shooting).

We conclude that the emotional reaction to terrorism and mass shooting events are similar in most respects. Our main observations are: there is indeed an emotion shift, where anger, fear and sadness are the most expressed emotions; the emotions expressed are directly related to gender (fear for women, and anger for men); compared to anger, fear is more related to younger populations; the influence of the location on the emotion could only be observed for mass shooting events; the influence of the number of casualties was inconclusive; tweeters expressing fear and sadness tend to regard themselves as potential victims, sharing words of affection and support, while tweeters expressing anger tend to use intense words of hate, intolerance and call for justice; keywords that identify event location, nature of events and emotions are commonly used.

With regard to related work, the striking differences of our work are:

- we address sentiment measured as basic emotions, while the majority of works address polarity, except (SINGH; CHOUDHARY, 2017);
- we applied deep learning techniques over an automatically generated set of training seeds, whereas works that employ supervised learning leverage feature engineering (MIRANI;

SASI, 2016; AZIZAN; AZIZ, 2017; GARG; GARG; RANGA, 2017; WANG; VARGHESE; DONNELLY, 2016);

- our analysis encompasses both the location and user demographics, whereas related works are limited to geo-referenced tweets (MANSOUR, 2018; WANG; VARGHESE; DONNELLY, 2016);
- we draw conclusions using twelve mass violent events of two different types. All works address either terrorism or mass shootings, and only sample a single event. The only exception is (JONES et al., 2016), which encompasses three case studies.

The remainder of this work is structured as follows. Chapter 2 provides a background on concepts and techniques used in this work. Chapter 3 describes related work. Chapter 4 describes the methods, materials and approaches used to gather our data, develop an emotion classifier in our context, and provide answers to our research questions. Chapter 5 describes our experiments to find a suitable model for emotion prediction. Chapter 6 describes the analysis performed over the data to answer our questions. Finally, Chapter 7 presents conclusions and opportunities for future work.

2 THEORETICAL BACKGROUND

This Chapter presents the theoretical background needed for understanding the proposed work. We present definitions, approaches and techniques related to sentiment analysis, covering machine learning and deep learning; and metrics that are used to evaluate our results.

2.1 Sentiment Analysis

Sentiment Analysis is the field of study that deals with the automatic extraction and interpretation of people's opinions, attitude, and emotions from texts. Sentiment analysis is a natural language processing (NLP) problem and thus involves all inherited issues (LIU, 2012). The most popular form of sentiment analysis is opinion mining, sometimes even used as a synonym. Opinion mining describes tasks that mainly focus on determining the *polarity* of the expressed sentiment towards a target, typically positive, negative, or neutral.

Sentiment analysis is applied in different levels (LIU, 2012; MOHAMMAD, 2016):

- Document Level: Tasks in this level aim at determining whether the opinion of a whole document is positive or negative, or if there is no opinion (commonly referred as neutral opinion). Document-level sentiment analysis assumes that the opinion of a document is expressed towards a single target, such as an organization, an event, or a topic. In various scenarios, the document may contain opinions towards different targets, and thus the task can be broken down into more granular components of the document, such as sentence, aspect or word level sentiment analysis.

- Sentence Level: Tasks in this level assume that the document may describe different targets, thus each sentence may potentially describe sentiment towards a distinct entity (or a particular aspect of an entity).

- Entity and Aspect level: Tasks in this level perform fine-grained analysis to determine sentiments on targets (entities and their aspects), as sentiments alone may have restricted use.

In the proposed work, we apply sentiment analysis at document level (tweets).

2.1.1 Emotion mining

Emotion mining is an area within the sentiment analysis domain that refers to tasks that aim at the identification of affect states according to an emotion model (MUNEZERO et al., 2014), thus extending traditional polarity classification. Emotions are subjective feelings/thoughts and can have different intensities (LIU, 2012). There are different emotion models, among them basic emotions (EKMAN; FRIESEN, 1982), emotion dimensions (RUSSELL; MEHRABIAN, 1977) and cognitive-appraisal categories (ORTONY; TURNER, 1990). The basic emotions model is the most deployed one in emotion mining due to the availability of resources (e.g. annotated corpora, sentiment lexicons) and facility in evaluating results.

There is no agreement towards a common set of basic emotions, and different sets of emotions have been proposed. Ekman's basic emotions (EKMAN; FRIESEN, 1982) is one of the most popular emotion model, composed of six emotions: hapiness, sadness, anger, fear, disgust, and surprise. Plutchik's set of basic emotions (EMOTION. . . , 1981) is also popular, and includes eight emotions: joy, sadness, anger, fear, disgust, surprise, trust, and anticipation.

In the proposed work, we focus on emotions in our sentiment analysis tasks, as we want to compare emotional reactions on tweets related to mass violent event. We adopt the Ekman's basic emotion model, due to its popularity and facility of interpreting results. However, we disregarded happiness in our research, as we assumed such an emotion is not likely to be expressed in reaction to such a type of event[1].

2.1.2 Sentiment Analysis on Twitter

The Twitter platform is a popular service of news and social networking. The nature of Twitter allows users to post real time opinions and thoughts on diverse topics, producing a huge volume of valuable information that can be exploited for different purposes in the field of sentiment analysis, such as to study and understand social phenomena (ZIMBRA et al., 2018; AGARWAL et al., 2011). Posts on Twitter are called tweets, limited length messages that users share freely. Data produced on Twitter is in constant growth, as opinions influence human behavior and are present in almost any activity and interaction. Much of the decisions taken by individuals and companies are based on other‘s opinions and evaluations, thus confirming the value of Twitter as a source of data for studies in sentiment analysis domains (ZHANG; WANG; LIU, 2018).

[1]https://www.paulekman.com/blog/our-emotional-reactions-terrorism/

Sentiment analysis on Twitter is difficult due to the limited length of the tweets, written in informal language, and the possibility of containing grammar errors, abbreviations, use of jargon and popular internet expressions such as emoticons. Due to their limited length, tweets also contain a problem known as data sparsity, in which users use a wide range of irregular terms to refer to the same thing.

In the proposed work, we focus specifically on sentiment analysis of Twitter data, addressing its peculiarities by disregarding specific textual features of tweets and applying deep learning over the (almost) raw data.

2.2 Approaches for Sentiment Analysis

There are two basic approaches for sentiment analysis: lexicon-based and machine learning. Each one of the approaches encompass different concepts and techniques, which will be explained separately.

2.2.1 Lexicon-based approaches

Lexicon based approaches consist on use of pre-tagged lexicons. Lexicons are collections of entries (typically words) that are associated with a sentiment measure. Lexicons are commonly referred to as dictionaries. In such an approach, the sentiment in a piece of text (e.g. document, sentence) is calculated based on some aggregation of the sentiment scores of the individual terms/phrases. Sentiment lexicons can be manually or automatically created (SEVERYN; MOSCHITTI, 2015a; DESAI; MEHTA, 2016; TABOADA et al., 2011):

- Manually created lexicons: they are based on human effort to compile word-sentiment associations of high quality. In manually created lexicons, negation (e.g., not good), intensification (e.g., very good) and sentiment-bearing words (e.g. adjectives and adverbs) are taken into account in the building process. A major advantage of manually created lexicons is that since they are built by humans, the correctness rate is higher. On the other hand, one drawback is that such a task is time consuming, thus generally restricting the size of the dictionary and yielding low recall results. Recall is a measure that indicates the number of relevant instances that are retrieved from the total set of instances. In the context of dictionaries, the problem of low recall occurs when dictionaries of reduced size are not able to provide enough information for properly asserting on the sentiment of a piece of text, thus not considering it as belonging to a given sentiment class.

- Automatically created lexicons: they rely on the expansion of an initial list of words with labels automatically inferred from textual representations, such as emoticons and hashtags in the case of tweets. One advantage of this approach is that it is not as time consuming as manually created lexicons, but that comes at the expense of accuracy. There are popular examples of automatically created lexicons, such as the SentiWordNet (BACCIANELLA; ESULI; SEBASTIANI, 2010) for polarity and the NRC (MOHAMMAD; TURNEY, 2013a) for emotions.

Due to the generally low recall and to the lack of dictionaries in our specific context, and also because in tweets many expressions are not contained in the dictionaries (e.g. abbreviations, informal language), we disregarded lexicon-based approaches in this work.

2.2.2 Machine Learning Approaches

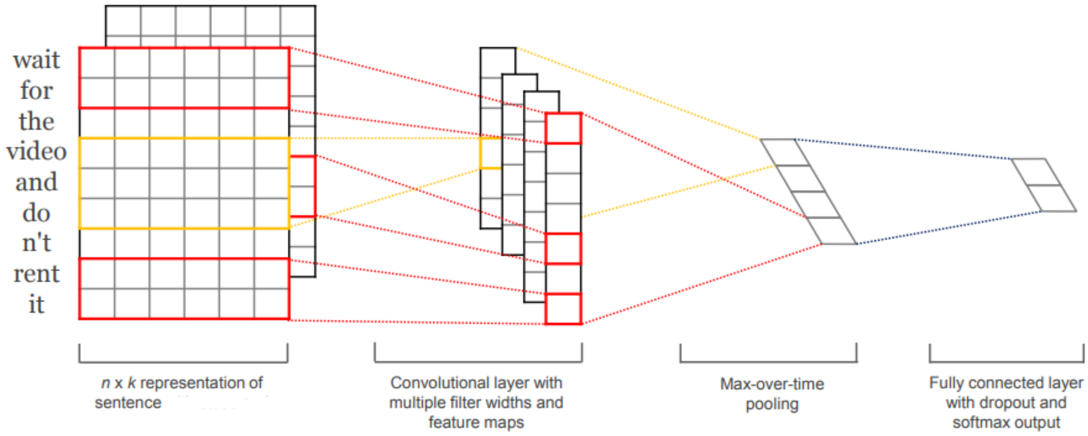
Machine learning refers to a set of methods that can automatically learn patterns from data and use them to predict future data, or to perform other kinds of decision making tasks (MURPHY, 2012). Inputs to machine learning methods are features extracted from pieces of text in any granularity, ranging from documents to sentences or clauses.

One of the most popular approaches in machine learning is the supervised learning, although others are possible (e.g. unsupervised learning). In supervised learning, algorithms are provided with engineered features and also with examples of associations, i.e., instances that are previously labelled, composing an annotated corpus. Typically in supervised learning algorithms the data is divided into train data (labelled instances used during the learning process) and test data (labelled instances used to evaluate the algorithm), and the algorithm has to learn how to label new unseen instances based on the acquired knowledge from the train data with help of the engineered features, generating then a prediction model (MURPHY, 2012; GOODFELLOW; BENGIO; COURVILLE, 2016; SHALEV-SHWARTZ; BEN-DAVID, 2014). Support Vector Machine (SVM), Naive Bayes (NB) and Maximum Entropy (ME) are popular examples of typical supervised learning algorithms.

Typical supervised learning algorithms rely on *feature engineering* and *feature selection* during the learning process. Feature engineering is the process of using domain-specific knowledge to extract representative information from raw data that help supervised learning algorithms to perform effectively. This process can be performed with help of domain-experts, that indicate which fields/attributes from the data are the most informative ones, or through "brute-force", where all the data is fully measured so that the relevant information can hopefully be isolated. Feature selection is the process of identifying and removing the irrelevant/redundant features, thus reducing dimensionality of the data and retaining the most representative features (KOTSIANTIS, 2007). Although qualified methods of feature engineering and feature selection help supervised learning algorithms in performing effectively, these processes are generally empirical, based on linguistic intuition, and trial and error. Also, they are task dependent, requiring additional research in each new task (COLLOBERT et al., 2011).

The quality and size of the annotated corpus is also crucial in training a supervised learning algorithm effectively. Context-specific annotated corpus is preferable, as sentences and words may have different meanings from one context to another (LIU, 2012; ZIMBRA et al., 2018). The manual creation of such a corpus is a costly activity, that also incorporates relatively little information, often insufficient to estimate relevant features in the feature engineering process (MURPHY, 2012). In addition, a corpus of insufficient size may raise problems such as *underfitting* and *overfitting*. The first problem occurs when the algorithm's prediction model is too simplistic to represent the relationship between the instances and their labels. On the other hand, overfitting occurs when the prediction model is so complex that the model fits to the dataset too closely, thus not working well on the presence of noise in the data (MURPHY, 2012; KELLEHER; NAMEE; D'ARCY, 2015). An alternative to the task of manually annotating a corpus is to automatically perform data labelling. Different approaches have been proposed, and leverage domain-specific features. For instance, in sentiment analysis, approaches that exploit emotion-word hashtags in tweets are deployed to automatically annotate corpora for sentiments (WANG et al., 2012; MOHAMMAD, 2012; MOHAMMAD; KIRITCHENKO; ZHU, 2013). They rely on an initial set of labelled instances (dictionary) (WANG et al., 2012), on emotion-words such as Ekman's basic emotions (MOHAMMAD, 2012) for filtering emotions in tweets, and on existing sentiment-words corpora for filtering sentiments (polarity) in tweets (MOHAMMAD; KIRITCHENKO; ZHU, 2013). Tweet-specific features are explored in the process, combining different machine learning approaches for expanding an existing corpus or creating new ones from the scratch. Another known approach to automatically label data is called *distant supervision*, in which already existing domain-specific corpora are used to automatically collect examples of relations in other contexts. Distant supervision has been explored in the last years (SUTTLES; IDE, 2013; GO; BHAYANI; HUANG, 2009a; PURVER; BATTERSBY, 2012; SEVERYN; MOSCHITTI, 2015b), and results suggested that it is an effective way to automatically perform data labelling.

Figure 2.1: Example of a CNN architecture in NLP



Source: Adapted from Kim (2014).

In recent years, alternatives to the time consuming and error prone tasks of both manual annotation and feature engineering have gained importance. A popular technique to deal with these restrictions of traditional supervised machine learning approaches is *deep learning*. Deep learning can perform well without the need of hand-crafted engineered features and large sets of labelled data. This is the technique we adopt in our work.

2.2.3 Deep Learning

Deep learning has emerged as a powerful technique that makes use of computing power and large sets of data. In recent years, it has shown great value in applications of NLP. Conceptually, it is a specific type of machine learning technique that applies *neural networks* for learning features at increasing levels of abstraction in multiple layers (MURPHY, 2012; ZHANG; WANG; LIU, 2018). The basic architecture of a deep learning model is comprised of multiple layers of processing units that are responsible for the extraction and transformation of features. Roughly, lower layers that deal with raw data learn simpler features, whereas higher layers learn more complex ones, derived from outputs of lower layers. The result is a set of features that hierarchically grows in complexity.

Deep learning has being applied to solve problems in contexts such as computer vision, speech recognition, and natural language processing, among others. Kinds of data used in deep learning models vary from one context to another, as well as the data handling process. For example, computer vision tasks require processing a large number of input features (pixels) from images that should be standardized in order to have their pixels values to fall into reasonable ranges. Another example is the document modeling in NLP tasks, so that sentences and words are converted into ranges of values (GOODFELLOW; BENGIO; COURVILLE, 2016). Specifically in NLP applications, deep learning models generally rely on *Word Embeddings* as input features.

2.2.3.1 Word Embedding

In NLP, Word Embedding is the result of the process of deriving distributed representations of words in a vector space. Through a set of techniques for language modelling and feature learning, words are mapped into vectors of numbers. Each vector represents the features of a specific word in the vocabulary, describing linguistic properties and patterns. Traditional approaches for generating word embeddings generally rely on neural networks or word-related matrix operations (LEBRET; COLLOBERT, 2014; MIKOLOV et al., 2013b). Popular models for generating word embeddings are the *Word2Vec* (MIKOLOV et al., 2013a) and the *GloVe* (PENNINGTON; SOCHER; MANNING, 2014). The latter model provides pre-trained word embeddings generated on top of a huge volume of Twitter data (about 2 billion tweets).

In this work, we use the GloVe‘s pre-trained word embeddings as one of the inputs for generating our emotion prediction model.

2.2.3.2 Neural Networks

Neural networks are networks composed by neurons connected and distributed in multiple layers. Neurons are units responsible for processing the information during the learning process. Neural Networks are able to adjust themselves, changing weight values in its connections based on the data. Such a property enables neural networks to learn how to perform, for instance, classification tasks (ZHANG; WANG; LIU, 2018; ZHANG, 2000). Zhang at al. (ZHANG; WANG; LIU, 2018) categorizes neural networks into two types:

- Recursive/Recurrent Neural Networks (RNN): this type of neural network has its connections between neurons forming a directed cycle and is able to apply a set of weights recursively over the input. This property gives RNNs the ability to have its neurons performing tasks over the input based on previous computations, thus working with a sort of internal memory (ZHANG; WANG; LIU, 2018; SCHMIDHUBER, 2015).

- Feedforward neural networks (FNN): differently from RNNs, this type of neural network does not have its connections between neurons forming a directed cycle. Basically, the information in a FNN flows only forward in only one direction: from the input nodes to the output ones, passing through intermediate nodes in the way (ZHANG; WANG; LIU, 2018; SCHMIDHUBER, 2015).

Deep Neural networks are different from traditional neural networks, as their structure are comprised of more layers and more units within a layer, thus being able to represent functions of increasing complexity. Deep neural networks are increasingly being used in deep learning models for NLP tasks, specially sentiment analysis. Examples include use of RNNs (SOCHER et al., 2013; LI et al., 2014), and special types of FNNs such as the CNNs (KIM, 2014; SEVERYN; MOSCHITTI, 2015b), among others (COLLOBERT et al., 2011). CNN models particularly have achieved good results in traditional NLP tasks, including classification tasks in sentiment analysis (KIM, 2014; SEVERYN; MOSCHITTI, 2015b; COLLOBERT et al., 2011; SHEN et al., 2014), and will be explained in details.

2.2.4 Convolutional Neural Networks

Convolutional Neural Networks are FNNs originally designed for computer vision that relies on multiple convolutional layers. Convolution is an operation that roughly takes in two inputs (two functions) to produce an output (a third function).

The basic architecture of a CNN is shown in Figure 2.1, which illustrates an application in NLP. There are four main stages in a typical CNN architecture (KIM, 2014; ZHANG; WANG; LIU, 2018; GOODFELLOW; BENGIO; COURVILLE, 2016):

- In the first stage we have the input, which are representations of sentences/words. As mentioned in Section 2.2.3, in NLP typically these representations are in the form of word embeddings.

- In the second stage we have a convolutional layer that performs *convolutions* over the input. A convolution operation basically applies filters of pre-defined sizes over the input to produce a new feature. Each input region in which a filter is applied on can be called a projected region. Filters are also known as kernels, and output features as *feature maps*. In convolutions, filters of different sizes slide all over the input, multiplying its weight values to the values of the projected regions. The values originated from such operations are summed up in one single value for each region, that represents it. The application of *N* different filters, producing *N* arrays, generates a stack of feature maps.

- In the third stage we have a layer responsible for a subsampling process (also known as *pooling*) that reduces the size of the feature maps. The goal is to reduce the number of features while retaining the most valuable ones (the ones with the highest values). A common pooling operation used is the *Max Pooling*.

- In the fourth stage we have a fully connected layer (which connects every layer to the other ones) responsible for regularization and normalization. The regularization is performed in order to prevent overfitting, and basically consists of adding/removing information that enables the network to learn useful features. A common deployed technique for regularization is the *dropout*, in which units of the network are dropped out and reinserted after training. Finally, the normalization is responsible for normalizing the outputs into probability distributions over the predicted output classes. A common used function is the *Softmax*.

An important part of CNNs, as one can note, are the convolutions performed in the convolutional layer. Convolutions are basically responsible for extracting local features. In NLP, this characteristic allows combinations/sequences of phrases/words to be discovered as good indicators of the topic of a document for example (ZHANG; WANG; LIU, 2018), or even as indicators of emotions in tweets. As mentioned earlier, works have deployed CNNs in NLP tasks, achieving good results (KIM, 2014; SEVERYN; MOSCHITTI, 2015b; COLLOBERT et al., 2011; SHEN et al., 2014). In special, Kim (2014) trained a Convolutional Neural Network on top of pre-trained word embeddings. Models were evaluated against several datasets and the results outperformed several state of the art methods in the majority of the experiments. Kim's network architecture has been widely referenced in related works. For instance, Severyn and Moschitti in (SEVERYN; MOSCHITTI, 2015b) describe a new model for initializing the network's parameters, showing that their result could be ranked top two in different tasks performed over the datasets provided by Semeval-2015 [2].

In this work, we use a CNN to build a deep learning model to predict emotions in our mass violent events tweets. Specifically, we adopted the CNN architecture of Kim (2014) due to its results, and pioneering.

[2]http://alt.qcri.org/semeval2015/

2.3 Evaluation Metrics

Performance evaluation of classification models rely commonly in four metrics (KHARDE; SONAWANE, 2016; WANG; YAO, 2012; HOSSIN; M.N, 2015; SOKOLOVA; LAPALME, 2009):

- Accuracy (Equation 2.1): accuracy measures the ratio of correct predictions over the total number of instances evaluated. Its focus is on the overall effectiveness of the classification model.
- Precision (Equation 2.2): precision measures the exactness of the classification model.
- Recall (Equation 2.3): recall measures the completeness of the classification model.
- F-measure (Equation 2.4): F-measure takes into account both precision and recall to basically express their tradeoff.

$$Accuracy = \frac{\sum_{i=1}^{c} \frac{TP_i+TN_i}{TP_i+TN_i+FP_i+FN_i}}{c} \tag{2.1}$$

$$Precision = \frac{\sum_{i=1}^{c} \frac{TP_i}{TP_i+FP_i}}{c} \tag{2.2}$$

$$Recall = \frac{\sum_{i=1}^{c} \frac{TP_i}{TP_i+FN_i}}{c} \tag{2.3}$$

$$F - Measure = \frac{(\beta^2 + 1) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{2.4}$$

Equations are expressed in macro-averaging, i.e., they treat all classes equally. For all equations, TP_i, TN_i, FP_i and FN_i have the following meaning:

- TP_i (True Positive): instances of the data that the classification model correctly predicted as belonging to a given class *i*;
- TN_i (True Negative): instances of the data that the classification model correctly predicted as not belonging to a given class *i*;
- FP_i (False Positive): instances of the data that the classification model wrongly predicted as belonging to a given class *i*;

- FN_i (False Negative): instances of the data that the classification model wrongly predicted as not belonging to a given class i.

All the above mentioned metrics are applicable to multiclass problems, and evaluate how well a classifier performs. In the proposed work, we use precision, recall and F-Measure to evaluate our prediction models.

3 RELATED WORK

In this Chapter, we examine related works on mass violent events on Twitter. We first present works that explore social aspects in Twitter in other contexts, showing common techniques and challenges in exploring Twitter data. Afterwards, we describe works focused specifically on mass violent events, and present information such as kinds of targets, classification algorithms, temporal window of analysis, kinds of data extracted from Twitter, and techniques deployed to gather them. All related works are summarized in one comparative table, in which we included this work in order to illustrate the differences.

3.1 Social Aspects in Twitter

Social aspects have been studied using Twitter data with different purposes, such as understanding gender-based violence (ELSHERIEF; BELDING; NGUYEN, 2017), estimating changes in population obesity rates (MITCHELL et al., 2013), investigating racial equality movements (CHOUDHURY et al., 2016), characterizing urban mobility and social relations patterns (GALLEGOS et al., 2016; LERMAN et al., 2016), among others. These works leverage sentiment analysis and incorporate information such as geo-location of the tweet and/or demographics of the users in their studies. Their results have shown the power of leveraging Twitter data to reach relevant findings.

All of the abovementioned works had to overcome challenges of data gathering to reach their goals. With regard to location, typically only a small fraction of tweets are geo-referenced. A common get around to this issue is to infer the tweet location from its user profile (SAKAKI; OKAZAKI; MATSUO, 2010; CHOUDHURY et al., 2016). Demographic information cannot be directly extracted from tweets, nor the respective profiles metadata. One alternative is to use census data of regions (such as cities and states) and associate them to geo-referenced tweets (GALLEGOS et al., 2016; LERMAN et al., 2016; MITCHELL et al., 2013). An alternative to estimate age and gender from the profile image using Face++[1] was proposed in (ELSHERIEF; BELDING; NGUYEN, 2017).

[1]https://www.faceplusplus.com/

3.2 Mass violent events in Twitter

Related work on mass violence context are described separately in two categories: the ones focusing on terrorism events, and the ones focusing on mass shooting ones. We address each one of the categories individually.

3.2.1 Related work on terrorism events

Works that leverage sentiment analysis techniques to understand terrorism focus mostly on post-event sentiments (BURNAP et al., 2014; GARG; GARG; RANGA, 2017; SIMON et al., 2014; CHONG, 2016) and on the perception of terrorist organizations (AZIZAN; AZIZ, 2017; MIRANI; SASI, 2016; MANSOUR, 2018). They all develop their respective analysis on collected tweets using keywords (hashtags, terms or urls) that describe the target (e.g. #prayForParis, ISIS).

Using sentiment analysis packages available in R, Chong (2016) analyzed tweets related to a series of coordinated terrorist attacks occurred in Paris in order to verify whether or not the emotion contagion theory (i.e. related people tend to have more similar sentiments) holds true. In his work, tweets were collected by filtering post event tweets, from the Twitter API, that contained a specific hashtag that related to the event (#prayForParis). The method used to measure quantitative sentiment score was based on the combination of pre existing R libraries. Content analysis was performed through topic extraction, achieved with the application of the SAS Enterprise Miner[2], a solution that creates predictive and descriptive models on large volumes of data. Word clouds related to the events tweets helped in identifying dominant and emerging topics. His results confirmed that the emotional contagion theory also applies to social media.

The influence of the sentiment in the post-event tweets information flow model is investigated in (BURNAP et al., 2014; GARG; GARG; RANGA, 2017). An information flow prediction model developed for tweets related to a terrorism act occurred in the UK was presented by Burnap et al. (2014). In their work, data collection was performed in a 14 days time frame, as it was the issue attention cycle to the event, that is, the period in which the public attention increased and then decreased. Tweets were gathered from the Twitter API, based on manual identification of trending keywords that were used to retain relevant tweets. Sentiment

[2]https://www.sas.com/en$_us/software/enterprise - miner.html$

polarity was measured using the existing dictionary-based tool SentiStrength[3]. Their results showed that emotive content is statistically significantly predictive of both size and survival of information flows of this nature. The authors also concluded that negative tweets largely outnumbers the positive ones, and that both timelags between retweets and the sentiment expressed are predictive of size and survival of information flows. Furthermore, offline content published on tweets is predictor of size, and tension is predictor of survival.

A model for tweets related to a terrorism event occurred in India was presented by Garg, Garg and Ranga (2017). Data collection was performed over a period of 1 month and only tweets containing event-related hashtags were retained. Features extracted from tweets (e.g. contents, retweets, favorites) as well as additional information such as last retweet time were used to study tweet survival. The sentiment classification algorithm used is based on an ensemble classifier developed by the authors that combines SVM and Naive Bayes. The classifier takes as input a log-ratio vector of positive/negative documents and relies on features such as bigrams and trigrams. In their approach, the number of retweets and the number of favorites worked as predictors of information flow and also information size. Their analysis revealed that negative tweets tend to survive more than the positive ones.

Another study investigated tweets related to a four day siege after a mall attack occurred in Kenya (SIMON et al., 2014) to assess how social media could contribute to post-event crisis management. Data collection was performed retaining tweets that contained relevant hashtags and those posted by Twitter accounts of managers and organizations involved in the event's response. The authors used the Alchemy API[4] to measure and compare the sentiment of tweets coming from different sources (hashtags, managers' and organizations' accounts). The authors concluded that managers in the field are more optimistic than the emergency operation centers.

Other studies investigate the sentiment towards terrorist organizations. Mansour (2018) collected geotagged ISIS-related tweets from different countries and measured sentiments in order to compare how people from Western and Eastern countries view ISIS. Data collection was performed through the Twitter API, filtering tweets that contained the specific hashtag *"ISIS"* and that the respective geocode attribute refereed to a country within a pre-stablished set of eigth countries, in which four of them represented the Western world and the another four represented the Eastern one. Sentiment score was calculated by loading the Opinion lexicon (LIU; HU; CHENG, 2005) dictionary for use by R packages. With no significant differences, results showed that all users view ISIS as a source of threat, using the same words when tweeting about it.

[3] http://sentistrength.wlv.ac.uk
[4] Currently, this API is part of IBM Watson system.

Mirani and Sasi (2016) assessed different sentiment classification models for tweets containing ISIS-specific hashtags. They trained five different algorithms on a dataset in which the polarity label was automatically assigned using the Opinion lexicon. Note that the Opinion lexicon is a dictionary originally created for the domain of web product reviews to compensate the limitations of generic sentiment lexicons in that specific domain (LIU; HU; CHENG, 2005), thus its application in contexts such as terrorism may not perform well (for example a bias can be introduced due to a different meaning of a same piece of text in different contexts). The algorithms used were the Support Vector Machine, Maximum Entropy, Random Forest, Bagging, and Decision Trees.

An approach for detecting pro-terrorism behavior of Twitter users is described by Azizan and Aziz (2017). The authors collected tweets containing terrorism-related keywords (e.g. jihrad), and for the users who posted these tweets, they also collected their previous posts. Sentiment was detected using a Naive Bayes classifier trained over an automatically labeled dataset through SentiWordNet. Then, they developed a model to detect the sentiment towards terrorism using historical tweets from the users.

In all the above mentioned works, sentiment is restricted to polarity, and in most of them, the focus is on post-event tweets. To the best of our knowledge, the only work that addresses emotions in violent mass events is the one presented by Singh and Choudhary (2017), which analyses tweets about a terrorism event that occurred in Barcelona. The authors collected tweets using specific hashtags, and identified basic emotions using the lexicon NRC, concluding that fear was the most frequent emotion, followed by anger. However, the authors pointed out that NRC was not suitable to their purpose, as it did not contain entries for many words and expressions.

3.2.2 Related work on mass shooting events

Fewer works address sentiment analysis in mass shooting events, and all of them are restricted to polarity (VARGAS et al., 2016; WANG; VARGHESE; DONNELLY, 2016; JONES et al., 2016). Through crowd sourcing, Vargas et al. (2016) compared overall and targeted sentiments in Twitter using tweets related to three crisis events: two of them related to a natural disaster and the other, to a school mass shooting. Data collection was performed by filtering specific keywords/prhases related to the events. Then, the authors manually selected a list of sentiment targets that were somehow involved in the events, such as the political figures, organizations, and celebrities. Sentiment labelling was achieved through a crowdsourcing service.

In their analysis, they noticed that sentiment regarding the mass shooting event contained more targeted sentiments, (i.e. sentiments expressed towards mentioned subjects), compared to natural disasters, which involved an overall expressed sentiment. This finding is confirmed by a study about a school shooting, in which the polarity of sentiment related to pro/anti-gun stance was identified (WANG; VARGHESE; DONNELLY, 2016). To achieve that, the authors trained different classification algorithms using a manually labeled dataset, which was collected using references to the shooting, as well as to guns in general. Geolocations of the tweets were used in order to show the differences of these sentiments across US states. Analysis reported that on the day of the shooting the rates of both pro-gun and anti-gun sentiments were at elevated level, and that pro-gun sentiment remains higher for longer than the anti-gun sentiment, which quickly falls instead.

Using three violent college shooting events, Jones et al. (2016) investigated whether or not Twitter could be a viable data source to study the emotional impact of mass violent events on communities. They collected pre and post event tweets in order to perform their analysis, and targeted tweets from users located within impacted communities. First, users were randomly selected from accounts likely to be followed by community residents only (e.g. radio station) and official college accounts. Then, for each event, the authors collected tweets at account level posted weeks before the event, and after its occurrence. In possession of these tweets, they examined pre and post-negative emotion expression alone, and in conjunction with the use of event-related words among the different targeted users in order to compare their reactions. Negative emotion was identified using a dictionary-based approach (LIWC[5]), and keywords manually selected as representative of the attack were used to filter the tweets explicitly related to the event. In their analysis, the authors observed an increase in negative emotions after the events and differences with regard to time frame, where use of event-related words and levels of negative emotion were at highest right after the events. In addition, their comparison of community and college control groups indicated that the latter is more emotionally impacted.

3.3 Final Remarks

Table 3.1 summarizes the key aspects of related works on mass violent events, and the difference with regards to ours. First, we address emotional reactions in terms of Ekman's basic emotions model, as in (SINGH; CHOUDHARY, 2017). However, to overcome the limitations of an emotion lexicon for tweets in general, and the mass violence domain in particular,

[5]www.liwc.net

Table 3.1: Summary of Related Work.

Related Work	Target	Sentiment	Tweet Collection - Post Event	Tweet Collection - Pre-event	Sentiment Classification	Trainning Set	Demographics	Location
[12]	Terrorism act	Polarity	Keywords		SentiStrengh		Data on population and gun ownership	Geolocated tweets
[13]	Terrorism act	Polarity	Keywords		ML (ensemble of Naive Bayes and SVM)	log-ratio vector of positive/ negative documents		
[14]	Terrorism act	Polarity	Keywords, Official profiles		Alchemy API			Keywords (in tweet text)
[15]	Terrorism act	Polarity	Keywords		R packages			
[16]	Any reference to terrorism	Polarity	Keywords	previous tweets of users who posted about terrorism	ML (Naive bayes)	Automatically labeled using SentiWordNet lexicon.		
[17]	Terrorism Organization	Polarity	Keywords		ML (SVM, Maximum Entropy and 3 others)	Automatically labeled using Opinion lexicon.		Geolocated tweets
[18]	Terrorism Organization	Polarity	Keywords		Lexicon (Opinion)			Geolocated tweets
[21]	Terrorism act	Emotion	Keywords		Lexicon (NRC)			
[8]	Mass shooting events	Polarity	user accounts following public community accounts	user accounts following public community accounts, keywords	LIWC		Community sizes to sample users	user accounts following public community accounts as users from that area
[19]	Mass shooting event	Polarity	keywords (event and guns)	keywords (guns)	ML (SVM, Maximum Entropy and 6 others)	Manually labelled dataset of 5000 tweets		Geolocated tweets
[20]	Crisis tweets	Polarity	Keywords		Manual (crowdsourcing)			
This work	Terrorism and Mass shooting acts	Emotion	Keywords	Keywords	Deep Learning (CNN)	Filter by keywords	Age/Gender (face++)	user's profile location

we adopt a supervised learning approach. Unlike other works that combined feature engineering and machine learning algorithms (MIRANI; SASI, 2016; AZIZAN; AZIZ, 2017; GARG; GARG; RANGA, 2017; WANG; VARGHESE; DONNELLY, 2016), we adopt a CNN deep learning architecture (KIM, 2014). To automatically label the training dataset, we filter the seeds from the collected tweets dataset using keywords, which yielded superior performance compared to the use of emotion lexicons to label the seeds. Another distinctive feature is the number and type of mass violence events, as we encompass both terrorism (4) and mass shooting (8) events. We analyze the sentiment using both pre/post event tweets, as in (JONES et al., 2016), but with a different goal, and therefore we do not focus on the tweets of a specific sampled group. Finally, we use information extracted from users' profiles to analyze emotions according to users' demographics, more specifically location, gender, and age, a novel aspect of sentiment analysis in this domain.

4 MATERIALS AND METHODS

This chapter presents the materials and methods used in order to gather the necessary data to run our experiments and perform our analysis. First we detail the construction of our datasets, describing the data collection process and the text pre processing actions. Next we describe the retrieval of location and demographic information from our tweets. Afterwards, the creation of our emotion gold standards are detailed. Finally, we present our emotion classification model.

4.1 Datasets

Our tweet datasets encompass tweets related to twelve mass violent events. Four terrorism datasets and eight mass shooting events. Information for these events are summarized respectively in Tables 1.1 and 1.2.

We decided to target only events that took place in the US and in the UK, due to a series of factors. First, we wanted to focus on English-written tweets in order to disregard differences due to idioms. In addition, there are significantly more tools for natural language processing, as well as crowdsourcing services available for the English language.

The number and representativeness of the events in each category was another factor. With regard to terrorism, both countries have witnessed terrorism attacks recently and are highly concerned with the matter. As for mass shootings, the US has consistently experienced a high number of these incidents over the last years. We chose well-known mass shooting events that happened within the last 3 years, a period in which the use of Twitter is comparable. All of these events and their aftermath drew a lot of attention from the world media, and their impact is not restricted to their respective local communities and surroundings. Thus, they allow for a comparison of different locations, based on a large volume of data that is likely to be produced. Finally, the chosen events have different motivations, targets, methods, and resulted in a different number of casualties. This variety allows a comparison on the emotional impact both in terms of properties of the act, as well as the scale of causalities (injuries/ deaths).

The method to create the datasets explored in this work is depicted in Figure 4.1, which is explained in the remaining of this Section.

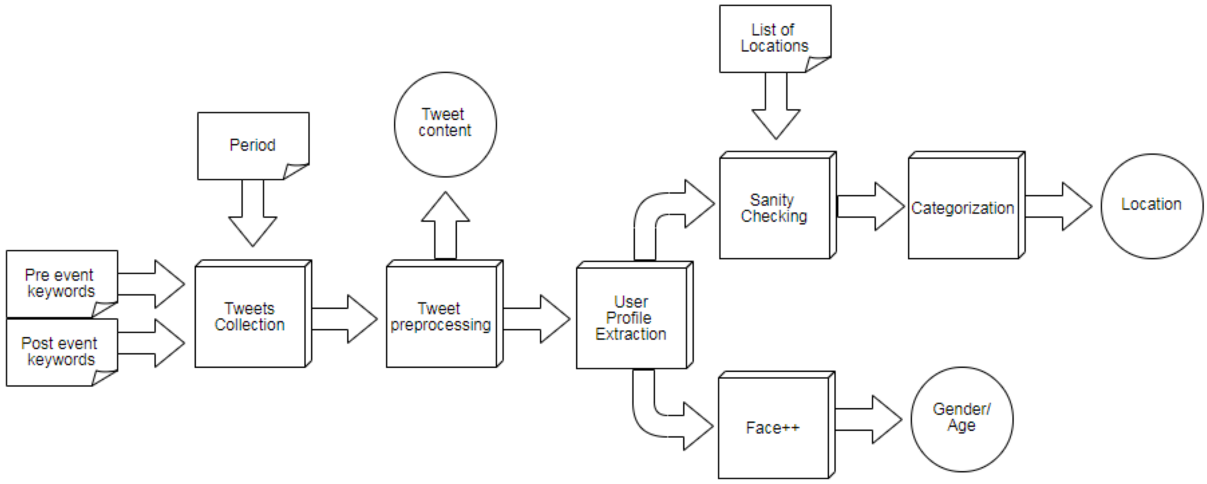


Figure 4.1: Process Diagram.
Source: The author.

4.1.1 Data Collection

Our research involves both pre and post-event tweets, in order to be able to investigate the emotional impact of events, and change with regard to a prior situation. As our goal is to understand the emotional reaction of the general population with regard to mass violence acts, the strategy of sampling users and collecting tweets from their accounts (JONES et al., 2016) was not suitable. Therefore, we collected both pre/post event tweets using keywords that characterized the community prior to the event, and the event itself.

For post event tweets, we manually inspected Twitter trending topics and raw data gathered from the web, as well as samples extracted using the official Twitter API on the respective event dates. We found recurrent hashtags for each one of the events and used them as query search terms.

To compose the sets of pre-event tweets, we queried tweets using the generic keywords referring to the locations where the events took place. We observed that these keywords were commonly used to tweet about local citizen's thoughts on diverse topics such as football teams, universities, and daily news regarding these locations. For terrorism events, we used the name of the cities where the acts occurred. For mass shooting incidents, as some communities were very small, we also used the name of the county and the name of the state.

For each event, we collected tweets two days before the event, the actual day it happened, and two days after the event, in a total of 5 days. We used an open source project[1] to overcome a restriction of the Twitter official streaming API, which does not allow the collection of an unlimited number of tweets from the past. The use of the API is straight-forward, relying

[1]https://github.com/Jefferson-Henrique/GetOldTweets-python

on the execution of a Python code along with a set of parameters, such as the user name to search tweets from a specific user, boundary dates to limit the search, and query texts to be matched in the search, among others. As parameters, we used query search terms combined with the respective boundary dates of the events. For example, one of the queries used to search tweets for the Manchester terrorism event was as follows: *"python Exporter.py –querysearch "#prayformanchester" –since 2017-05-22 –until 2017-05-24"*. Further details on the API can be found in the project's repository[2]. Tables 4.1 and 4.2 provide details on the parameters used to collect pre/post event tweets, as well as the respective number of resulting tweets, divided into pre-event (BEFORE) and post-event (AFTER) sets.

4.1.2 Pre-processing

We performed typical data pre-processing actions of tweets' text, removing hyperlinks, mentions (@), hashtags (#), other special symbols (&, /), and duplicated tweets (LIU, 2012). We also applied an English dictionary[3] to discard tweets with too many misspellings or tweets in other languages. We considered valid those tweets in which the pre processed text contained at least three words and that words were contained in the dictionary. Both pre-processing actions and the validation against the dictionary were performed by a Java[4] code.

As a result we produced a total of eight datasets for terrorism and sixteen for mass shooting events, divided into BEFORE and AFTER subsets. These datasets are available in a public repository[5]. The structures of these datasets are identical and include the fields listed below:

- Tweet ID
- Username
- Tweet text
- Date and Geolocation
- Number of retweets and favorites
- Mentions and Hashtags

Table 4.1: Query Terms, Dates and Dataset per Terrorism Event

Event Name	Query terms	Period	BEFORE (#tweets)	AFTER (#tweets)
#prayformanchester	#prayformanchester "manchester"	05-20-2017 to 05-24-2017	BM (5,351)	AM (25,010)
#londonbridge	#LondonBridge "London"	06-01-2017 to 06-05-2017	BL (20,379)	AL(29,656)
NY-october	#NYCStrong, #manhattanattack "new york"	10-29-2017 to 11-02-2017	BNYO (12,130)	ANYO (11,072)
NY-december	#nycexplosion #nycbombing "new york"	12-09-2017 to 12-13-2017	BNYD (64,906)	ANYD (3,602)

Table 4.2: Query Terms, Dates and Dataset per Mass Shooting Event

Event Name	Query terms	Period	BEFORE (#tweets)	AFTER (#tweets)
lasvegas	#LasvegasShooting,#VegasStrong #VegasShooting,"las vegas","nevada"	09-29-2017 to 10-03-2017	B (5,467)	A (59,871)
marshallcounty	#KentuckyShooting,#MarshallStrong #KentuckySchoolShooting,#KentuckyStrong "marshall county","kentucky"	06-01-2017 to 06-05-2017	B (6,483)	A (3,025)
maryland	#CapitalGazzette,#AnnapolisShooting #CapitalGazetteStrong,#capitalgazetteshooting "annapolis","maryland"	06-26-2018 to 06-30-2018	B (4,722)	A (6,645)
orlando	#orlandoshooting,#pulseshooting #OrlandoStrong,"orlando","florida"	06-10-2016 to 06-14-2016	B (7,262)	A (56,936)
ranchotehama	#ranchotehamashooting,#tehamashooting #RanchoTehama,"rancho tehama","california"	11-12-2017 to 11-16-2017	B (9,670)	A (2,765)
santafe	#SantaFeShooting,#SantaFeStrong #prayfortexas,#prayforsantafe, "santa fe","texas"	05-16-2018 to 05-20-2018	B (16,054)	A (7,396)
stoneman	#ParklandSchoolShooting,#ParklandStrong #StonemanShooting,"parkland","florida"	02-12-2018 to 02-16-2018	B (13,468)	A (18,391)
sutherland	#texasshooting,#TexasStrong,#prayersfortexas "sutherland springs","texas"	11-03-2017 to 11-07-2017	B (13,781)	A (9,168)

4.2 Demographics and Location Extraction

We aimed to perform our analysis using the location of the tweet, together with the gender and age of the Twitter user. To that purpose, for each collected tweet, we searched for the respective user profile, from which we extracted such information. To search for the user profile, we used a Python API called Twython[6], which is a wrapper for the official Twitter API methods. The method used in order to obtain the needed information was the *"show_status"* one, which takes as input the tweet id and returns the complete tweet structure in *JSON*[7] format.

The whole tweet structure includes a sub-structure that provides information on the user

[2]https://github.com/Jefferson-Henrique/GetOldTweets-python
[3]https://github.com/dwyl/english-words
[4]https://go.java/index.html
[5] https://github.com/jonathasgabriel/Terrorism-Mass-Shooting-Twitter-Dataset
[6]https://twython.readthedocs.io/en/latest/
[7]https://www.json.org/

profile. However, the profile of a Twitter user does not contain age nor gender information, therefore, we extracted the user profile image and used the Face++[8] tool to obtain this information, as proposed by ElSherief, Belding and Nguyen (2017). Face++ takes as input images and outputs estimation of face attributes, including age and gender. Experiments performed using this tool (FAN et al., 2014) report an accuracy of 85%. The service provided by the tool was called through a Python code.

With regard to location, we observed that less than 1% of the collected tweets were georeferenced. We adopted the location as informed in the user's profile (SAKAKI; OKAZAKI; MATSUO, 2010). However, the location in each user profile is a free text informed by the user. For sanity checking, we compared each declared location against a list of cities in these countries[9][10]. Locations not matching any city in our list fall into the "other locations" category. The matching between the location informed by the user and the lists of cities was performed by a Java code.

Our original plan was to analyze the sentiment in the local community, i.e. the city where the event happened, against the sentiment in other locations. However, the number of tweets for comparison in that granularity was very small. Alternatively, a representative number of tweets could be taken into account if different locations were grouped into a higher level abstraction. In this way, we would be able to analyze emotions from people closer to the impacted community, who possibly are more related to the victims or are victims themselves, as well as users outside these communities, who expressed their emotions as a response to the aftermath publicized on news or social media.

For the terrorism events, we applied the following categories: a) for the UK events, we considered three categories, namely, cities in UK, cities in US, and other locations; b) for US events, which took place in New York City, we considered cities within the state of New York, cities in US outside the state of New York, and other locations.

Regarding mass shooting events, we considered three categories: local, US locations other than local, and other locations. The local category was composed of the city (e.g. Las Vegas) or county (e.g. Marshall County) where the event took place, as well as of cities within the same state (e.g. Kentucky). This strategy aims to compensate the lack of tweets for events that occurred in very small communities.

[8]https://www.faceplusplus.com/
[9]https://github.com/grammakov/USA-cities-and-states
[10]https://www.paulstenning.com/uk-towns-and-counties-list/

4.3 Emotion Gold Standard

Our work focuses on five out of the the six basic emotion categories defined by Eckman (EKMAN; FRIESEN, 1982), namely anger, fear, sadness, surprise, and disgust. We focused on negative emotions only, because we assumed that people are not likely to express happiness in reaction to such kinds of events[11]. Our approach considers that a given tweet is related to one (and only one) of the emotion categories, regarded as the *predominant emotion*. To test the emotion classifiers proposed in our research, we built two gold standards, one for each type of event.

4.3.1 Terrorism Gold Standard

The Gold Standard for the terrorism dataset was produced using Amazon Mechanical Turk (AMT)[12]. We adopted the 5 aforementioned emotion categories, plus an extra "none" category. The set of tweets labelled using AMT were taken from the terrorism events that occurred in the UK. The author annotated 967 tweets that were likely to be in each of the categories due to the presence of emotion keywords and expressions. Emotion keywords were defined by data sampling. For example, the tweet "*Deeply saddened by the loss of 22 beautiful lives. we should not live like this.*" was labelled as sadness due to the expression "deeply saddened". On the other hand, the tweet "*It's so scary to not feel safe in this World*" was labelled as fear due to the expression "It's so scary", and so on. In the annotation process, the author distributed the tweets as evenly as possible into the 6 categories.

Afterwards, we created a Human Intelligence Task (HIT) with these tweets, where AMT annotators were asked to determine which emotion best described a tweet. We instructed annotators to choose the primary emotion if more than one emotion could be identified, and to choose "none" if no emotion could be clearly determined. We targeted the HIT to two master annotators, so that we would have three annotators in total, including the author. We retained all tweets with at least two annotation agreements, used as labels. The resulting gold standard, referred to as *GSTR*, is composed of 607 tweets (Table 4.4).

[11] https://www.paulekman.com/blog/our-emotional-reactions-terrorism/
[12] https://www.mturk.com/

Table 4.3: Tweet examples for both Terrorism (TR) and Mass Shooting (MS) Events

Emotion	Tweet
Anger	(TR) No but seriously I'm so pissed off at these bigots who think terrorism is an excuse for racism. (MS) Woke up pissed off because of this homophobia bullshit
Disgust	(TR) I feel so so disgusted by the events that happened to England today and 2 weeks ago. My heart is with you guys! (MS)The more press I read about the more disgusted I am of the world we live in and how another human can cause such devastation
Fear	(TR) What is happening to this world... I have never been more scared in my life... (MS) I'm so scared to go to concerts now..
Sadness	(TR) Devastating to hear of the tragic events at last night. Truly saddening. Our thoughts with all those families affected. (MS) Woke up to the saddening news from Las Vegas my prayers are with everyone and their families.
Surprise	(TR) I am truly just shocked by how evil mankind is. (MS) I'm shocked at what happened in Vegas. Now the worst mass shooting in US history
None	(TR) The best way to destroy an evil idea is to make it conceptually obsolete. (MS) Nothing more than a single shot weapon. I have always defended the 2nd amendment but last night changed my mind

Table 4.4: Gold Standards: Number of labelled tweets per category

Gold Standard	Anger	Disgust	Fear	Sadness	Surprise	None
Terrorism (GSTR)	82	116	85	179	71	74
Mass Shooting (GSMS)	33	28	18	36	13	27

4.3.2 Mass Shooting Gold Standard

The creation of our mass shooting gold standard was achieved in a different way. Our initial action was to create it as we did for the terrorism one, using AMT. However, the policies of AMT made it totally unavailable to Brazilian users, and other attempts to use crowd sourcing were not successful. Thus, we gathered some volunteers to label a smaller dataset according to a different method.

In sampling tweets from either terrorism events or mass shootings, we observed many similarities (i.e. common use of keywords) as illustrated in Table 4.3. For this reason we first classified our mass shooting tweets using a prediction model trained using tweets from both mass shooting and terrorism datasets (details in Chapter 5), and then randomly sampled 45 labelled tweets of each category, totaling 270 tweets. This set of tweets, without the predicted label, was given to three annotators (including the author), in the same format as in an AMT task, and using the same instructions. We retained for the gold standard only tweets with at least two agreements. The resulting set, referred to as *GSMS*, is composed of 155 tweets (Table 4.4).

4.4 Emotion Classification

In order to classify our collection of tweets, we applied deep learning by training a Convolutional Neural Network as defined in (KIM, 2014). Our choice is due to its results, and the pioneering in using such approach for classifying natural language. This model outperformed traditional methods, such as Support Vector Machine, in a variety of text classification tasks and since then it is widely referenced in the literature. Another motivating factor for using such an approach was the automatic learning capability that deep learning has by incorporating improved learning procedures that make use of computing power and training data, working well on large sets of data (AIN et al., 2017).

The CNN architecture is mainly comprised of four layers (as detailed in Section 2.2.4). The Python code of the CNN implementation we used is publicly available[13],[14], and it is designed on top of TensorFlow[15], an open source machine learning framework for high performance numerical computation that comes with strong support for deep learning. The actual structure of the network is depicted in Figure 4.2 and is explained in the remaining of this Section.

This CNN has two inputs. The main input is a word-level training set, namely *World-level Input*, which was automatically labeled from the set of collected tweets, given the lack of an annotated corpus in our domain. Our approach to create the training sets is described in Chapter 5. The second input is a set of pre-trained word embeddings, namely *Glove's Embeddings*. As suggested in (KIM, 2014), it is a means of improving performance when the training set is not large enough, given that deep learning benefits from large sets of data. We chose the word embeddings corpus provided by GloVe[16] because it is trained specifically from a huge corpus of tweets. All the other parts of the network structure represent operations performed in Tensorflow. In the case of our CNN, all operations are based on functions of the *nn*[17] package, which is a wrapper for neural networks operations. Four layers compose the CNN:

- First Layer: this layer is represented by the *Embedding* object, and is responsible for converting our world-level input into *word embeddings*. This is achieved with the *embedding_lookup* function. This function takes as reference the Glove's word embeddings corpus, and basically it will lookup embeddings for our world-level input. The output

[13]https://github.com/cahya-wirawan/cnn-text-classification-tf
[14]https://github.com/dennybritz/cnn-text-classification-tf
[15]https://www.tensorflow.org/
[16]https://nlp.stanford.edu/projects/glove/
[17]https://www.tensorflow.org/api$_{d}ocs/python/tf/nn$

contains all the embeddings for all words in the vocabulary.

- Second Layer: this layer is represented by the objects *Conv(3)*, *Conv(4)*, and *Conv(5)*, and is responsible for applying convolutions with multiple filter sizes over the word embeddings. For each filter size (3, 4, and 5), a convolution is performed with the *conv2d* function and a feature map describing the sentences is generated.

- Third Layer: this layer is represented by the objects *Global Pool* and *Concat*, and is responsible for filtering the most important features from the feature maps. This is achieved by performing a max pooling operation with the *max_pool* function. This function is applied for each feature map resulted from the second layer. Afterwards, the *concat* function is applied to merge the results into one big feature vector.

- Fourth Layer: this layer is represented by the objects *Dropout* and *Out*, and are responsible for regularizing and normalizing the CNN, respectively. Regularization is achieved with the *dropout* function, which basically avoids overfitting and forces the network to learn features that are individually useful. Finally, normalization is performed with Softmax through the *softmax_cross_entropy_with_logits* function, which normalizes the outputs into probability distributions over the predicted output classes.

4.5 Automatic Generation of Training Seeds

There is no publicly available corpus for the domain addressed in this work, and manual annotation is a costly and error-prone activity. Also, our limited number of labelled tweets would not provide a large set of data for properly training the CNN. Therefore, we tried different approaches for gathering enough training seeds for our emotion categories:

- Distant supervision (GO; BHAYANI; HUANG, 2009b; PURVER; BATTERSBY, 2012; SUTTLES; IDE, 2013): we applied distant supervision and used the emotion-labelled electoral tweets provided by (MOHAMMAD et al., 2015) as training seeds.

- Filtering by keywords: The process for obtaining our set of keywords was the same as for labeling tweets for our gold standard. To define the keywords, we analyzed samples of our datasets in search of keywords that were likely to represent emotions in a tweet. Random sets of tweets were selected in the process. After defining a set of keywords, we

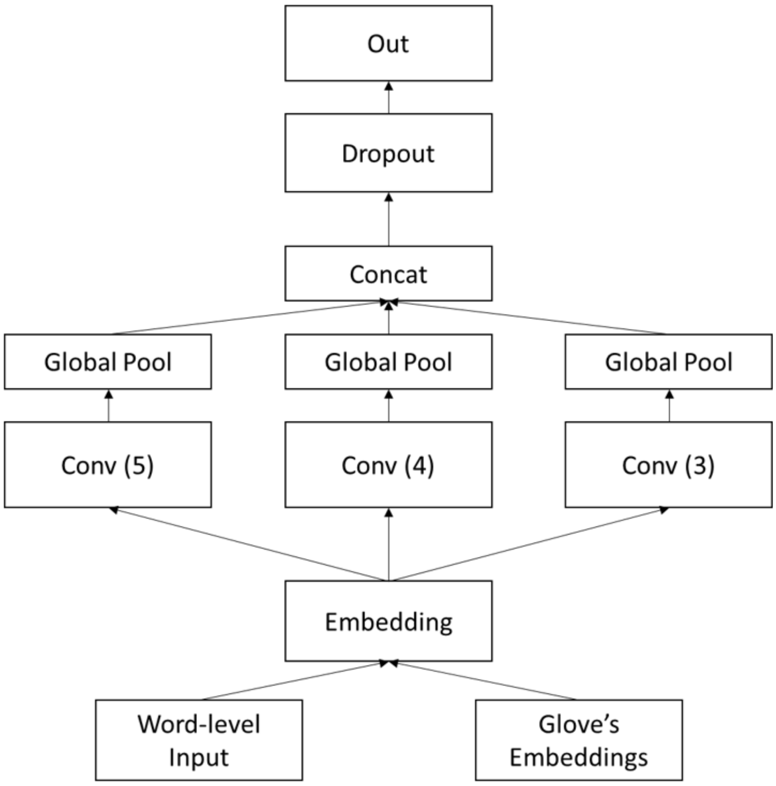


Figure 4.2: Convolutional Neural Network Architecture.
Source: The author.

sampled tweets containing such keywords and assessed that they were likely to belong to their respective emotion category. We also double-checked that these keywords were also represented in our gold standards. The set of keywords is shown in Table 5.3, which was used to filter seeds both for terrorism and mass shooting datasets. Tweets were retrieved from their respective post-event subset of tweets (AFTER). The resulting training sets are summarized in Table 5.4. The imbalance of the training sets are due to the number of collected post-event tweets (see Tables 4.1 and 4.2).

- Filtering by hashtags (MOHAMMAD, 2012): we used emotion hashtags collected from (MOHAMMAD, 2012) to provide automatic labelling. Labelled tweets were used as seeds.

- Dictionary-based filtering: we used a lexicon approach and filtered tweets with the emotion categories available in NRC (MOHAMMAD; TURNEY, 2013b). Tweets in which one emotion prevailed were filtered and used as training seeds.

In all of our experiments, training seeds for the "none" category were chosen by selecting tweets that did not contain any of the following terms: a) defined keywords used as seeds (Table 5.3), b) emotional hashtags defined in (MOHAMMAD, 2012), and c) emotion expressions labeled according to the NRC lexicon (MOHAMMAD; TURNEY, 2013b). As the number of retained tweets were much higher than the number for the other categories, we randomly selected a subset of tweets in which the size was the averaged number of seeds for the other five emotion categories.

5 MODEL EVALUATION EXPERIMENTS

In this chapter we describe the model evaluation experiments performed in order to generate our emotion prediction model. They are categorized in two types: experiments to firstly choose the model for automatic generating training seeds for our CNN, and evaluation experiments to find the most suitable emotion classification model for each kind of mass violent event. In all experiments, the CNN parameters we used were the same as in (KIM, 2014) because their results were built using these parameters, and all the variations we tried did not provide significant difference on our results. Furthermore, to improve our set of input data, we did as in (KIM, 2014) and loaded in our CNN pre-trained word embeddings for all the experiments. We chose the word embeddings corpus provided by GloVe [1] because it is extracted specifically from tweets. Following (KIM, 2014), the use of pre-trained word embeddings is an approach commonly used to improve performance when the training set is not large enough.

5.1 Experiments for Automatic Generation of Training Seeds

For each of the approaches described in Section 4.5, the CNN was trained and a prediction model was generated. The goal was to compare the performance of the different strategies to choose the one that performed better, in order to provide a representative set of seeds. The test was always conducted against our labelled set of terrorism tweets. This is due to the prior work we developed only for terrorism events (HARB; BECKER, 2018). General results of our models can be found in Table 5.1.

As we can see, distant supervision did not provide the best of the results. One explanation could be due to the peculiarities of our context, which includes words and expressions different than those of an electoral debate context. The approaches based on lexicon and hashtags did not provide good results as well. We noticed a very high level of absence of emotion hashtags in our dataset, which resulted in very few seeds, not enough to generate an accurate prediction model.

From all of our experiments, the one filtering by keywords provided the best results and therefore was the one used to generate our prediction model. We considered our model reliable because it achieved average precision and recall above 70%, which we believe were good results taking into account results presented in (SUTTLES; IDE, 2013; PURVER; BATTERSBY,

[1] https://nlp.stanford.edu/projects/glove/

Table 5.1: Results for the generated CNN prediction models

Approach	Avg. Precision	Avg. Recall	Avg. F-measure
Distant Supervision	0,4049	0,2099	0,1087
Keywords	**0,7348**	**0,7180**	**0,6846**
Hashtags	0,3220	0,3014	0,2783
Dictionary-based	0,5666	0,2958	0,2722

Table 5.2: F-measure for the model generated by filtering keywords

Emotion	anger	disgust	fear	sadness	surprise	none
F-measure	0,8603	0,6589	0,6280	0,7207	0,5932	0,6462

Table 5.3: Keywords used for filtering training seeds for the CNN

Emotion	Keywords
anger	anger, fuck, fucked, pissed, lmaof, damm
disgust	disgust, disgusted, disgusting
fear	worried, worry, scary, scaring, scared, fear
sadness	sad, sadness, saddened
surprise	surprised, surprising, surprise, shocked, shocking

Table 5.4: Number of training seeds per emotion category per context

Context	Anger	Disgust	Fear	Sadness	Surprise	None
Terrorism (TR)	672	300	854	2.007	186	803
Mass Shooting (MS)	2.973	695	2.058	3.127	2374	2.245

2012; MOHAMMAD et al., 2015). F-measure results for such a model with regard to each emotion category can be seen in Table 5.2. It can be seen that the model's result for anger stands out along with the one for sadness. Remaining emotions have similar results, excluding surprise that performed below 60% but still close to the average.

Table 5.5: Precision results for the emotion classification models

Emotion	anger	disgust	fear	sadness	surprise	none	macro-avg.
TR (GSTR)	0.79	**1.00**	0.53	**0.86**	**0.74**	**0.47**	0.73
TR+MS (GSTR)	**0.85**	0.98	**0.82**	0.84	0.68	0.31	**0.75**
MS (GSMS)	**0.88**	0.78	0.77	**0.82**	0.50	0.58	0.72
TR+MS (GSMS)	0.78	**0.80**	**1.0**	0.73	**0.59**	**0.65**	**0.76**

Table 5.6: Recall results for the emotion classification models

Emotion	anger	disgust	fear	sadness	surprise	none	macro-avg.
TR (GSTR)	**0.93**	**0.49**	**0.76**	**0.62**	**0.49**	**1.00**	**0.71**
TR+MS (GSTR)	0.93	0.46	0.60	0.59	0.43	0.98	0.67
MS (GSMS)	0.66	**0.89**	**0.77**	0.63	**0.69**	0.74	**0.73**
TR+MS (GSMS)	**0.82**	0.75	0.62	**0.87**	0.52	**0.77**	0.72

Table 5.7: F-Measure results for the emotion classification models

Model	anger	disgust	fear	sadness	surprise	none	macro-avg.
TR (GSTR)	0.86	**0.65**	0.62	**0.72**	**0.59**	**0.64**	**0.68**
TR+MS (GSTR)	**0.89**	0.63	**0.69**	0.70	0.53	0.48	0.65
MS (GSMS)	0.75	**0.83**	**0.77**	0.71	**0.58**	0.65	0.72
TR+MS (GSMS)	**0.80**	0.77	0.76	**0.80**	0.55	**0.71**	**0.73**

5.2 Experiments for Emotion Classification

We developed experiments with our CNN to find the most suitable classification model for our emotion categories. To choose the best emotion prediction models for each mass violent context, our experiments investigated whether or not the similarities of the respective tweets could be explored to improve the performance of the models, through a larger training set. Thus, three different models were trained, which varied in terms of the source of input training seeds: a) seeds extracted from terrorism-related tweets only (TR); b) seeds extracted from mass shooting-related tweets only (MS); and c) seeds from both events (TR+MS).

To asses each model, we used the respective gold standards, i.e. TR against GSTR, MS against GSMS and TR+MS against each gold standard individually. The comparison of the three models in terms of Precision, Recall and F-measure is displayed in tables 5.5, 5.6 and 5.7, respectively. In applying the emotion prediction models described in Section 4.5 to our tweet datasets, we noted that the prevalent emotions in terrorism tweets are anger, fear and sadness. This prevalence was also confirmed for the mass shooting ones. Thus, the evaluation of the models was influenced by these three emotions and the *None* category, which are essential to our analysis.

For the terrorism events, the performance assessed using GSTR is comparable for both models (TR and TR+MS), except for the category *None*, which is unacceptable (0.48%) for the model trained with TR+MS. This is due to a reduction of 16 percentage points (pp) in precision, and a similar recall ($\tilde{9}$8%). The performance for *fear* improved 7pp, at the expense of a drastic reduction in recall (16pp). Considering this performance, we decided to analyze the tweets related to terrorism using the model trained with TR seeds only.

For mass shooting tweets, assessed using GSMS, the model trained using TR+MS outperformed the one using MS in 3 out of these four categories. The differences are significant: 5pp for *anger*, 9pp for *sadness* and 6pp for *none*. These improvements are mostly related to recall. For *fear*, the results are comparable, but there is a trade-off between precision and recall: it achieved 100% in precision, at the expense of a 15pp reduction in recall. Thus, to classify tweets of the mass shooting datasets we used the model trained using TR+MS. Note that these results may be influenced by two factors: a) the GSMS gold standard may contain a bias, as the manually annotated instances were generated by a model trained on top of seeds from both contexts, and b) it is sensitive to the smaller size of the gold standard.

5.3 Concluding Remarks

In this Chapter we described the experiments we conducted in order to:

- choose which strategy to use for automatic generating training seeds for our CNN: the filtering by keywords strategy performed better and was the chosen one.
- choose which CNN model to use for each kind of event, where differences from model to model relied on the source of the training seeds: the chosen model for the terrorism events was the one trained with seeds filtered from terrorism events only, whereas for the mass shooting events the chosen model was the one trained with seeds filtered from both terrorism and mass shooting events.

6 ANALYSIS

In this Chapter we describe the analysis performed on our data in order to answer our research questions.

6.1 Q1: Is there an emotion shift due to mass violent events?

To answer this question, we analyzed the difference between the percentage of tweets with emotions in post/pre event tweets. We first present our findings for the terrorism events, and then for the mass shooting incidents.

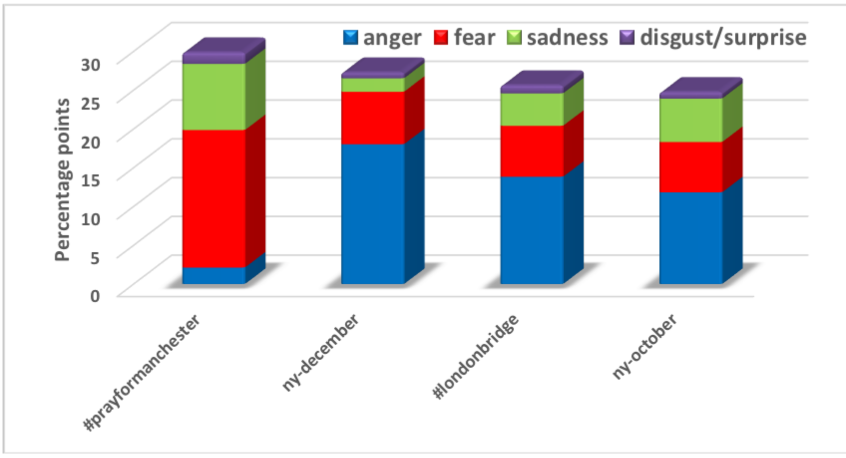


Figure 6.1: Increase in tweets with emotions after terrorism events.
Source: The author.

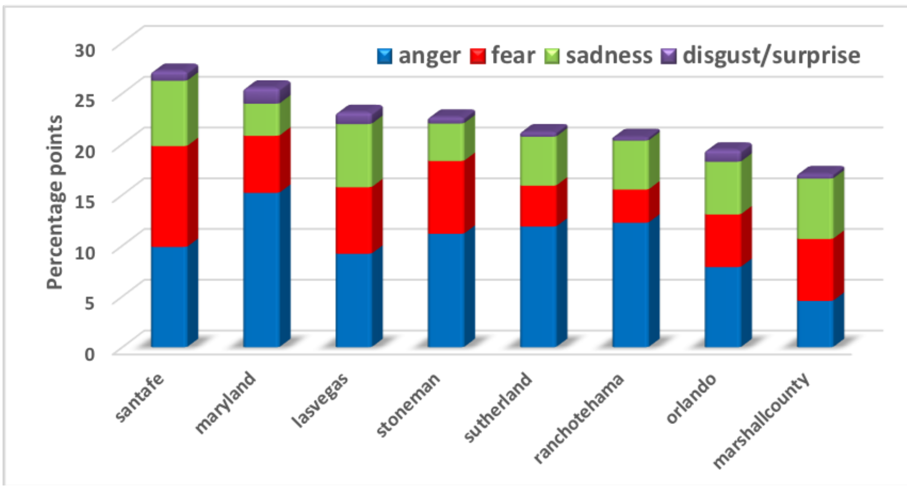


Figure 6.2: Increase in tweets with emotions after mass shooting events.
Source: The author.

Figure 6.1 depicts, for each terrorism event, the difference between the percentage of tweets with emotions in post/pre event tweets. On average, only about 7% of the tweets that preceded the events contained emotions, compared to an average of 32% after them. The contribution of each emotion to the difference is also represented in Figure 6.1. Three emotions prevailed after the events: anger, fear, and sadness. Anger was the most prevalent one in all events, except for the one in Manchester. In that event, a concert of a teen pop star, fear was the predominant emotion, followed by sadness. Differences were also observed for disgust and surprise, but these emotions are proportionally insignificant.

These same patterns were observed for mass shooting events. Figure 6.2 depicts, for each event, the difference between the percentage of tweets with emotions in post/pre event tweets. The increases range from 17 pp (Marshall County) to 27 pp (Santa Fe). We did not observe a relation between the target of the mass shooting (e.g. school, nightclub, church) and the increase.

We also observed the same prevalence of emotions in mass shooting events. In all incidents, anger was the prevalent emotion, followed by fear and sadness. The number of tweets associated with disgust and surprise did increase after the incidents, but they are proportionally insignificant. Figure 6.2 displays the contribution of these emotions to the observed difference. The biggest increase in anger was observed for Maryland, an event that targeted offices of a newspaper. For fear and sadness, the biggest increase was related to Santa Fe, an event that involved a school.

We conclude that there is indeed an emotional shift due to mass violent events in general, and that specifically anger, fear and sadness are the emotions that prevail once such events happen. Increases in surprise/disgust are also observed, but they are proportionally irrelevant.

6.2 Q2: Do different mass violent events evoke the same emotional reaction?

To answer this question, we compared the distribution of emotions for each event using post-event tweets. Figure 6.3 depicts this comparison for terrorism events, and Figure 6.4 for mass shooting incidents. In these graphs, the Y axis represents, for each event, the percentage of its total number of tweets distributed into emotion categories. Only tweets related to emotions are displayed.

Our analysis on terrorism events revealed that there are differences in terms of emotion distribution per event. As mentioned, anger is the most evoked emotion, followed by fear and

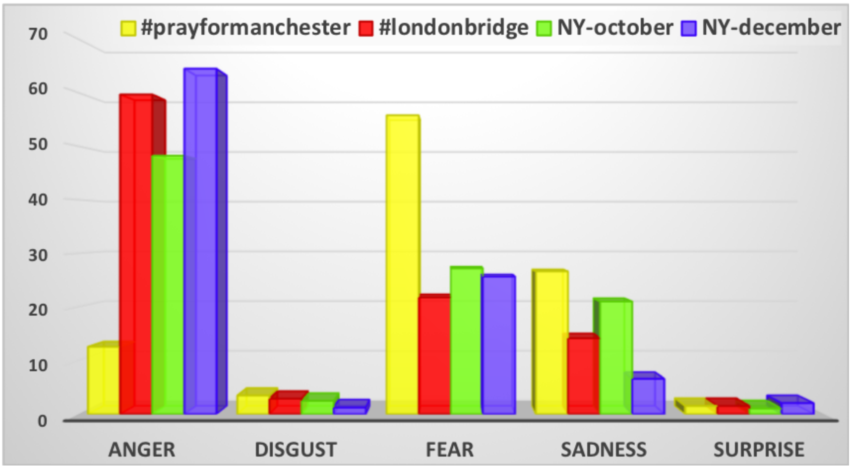


Figure 6.3: Emotion distribution for terrorism events.
Source: The author.

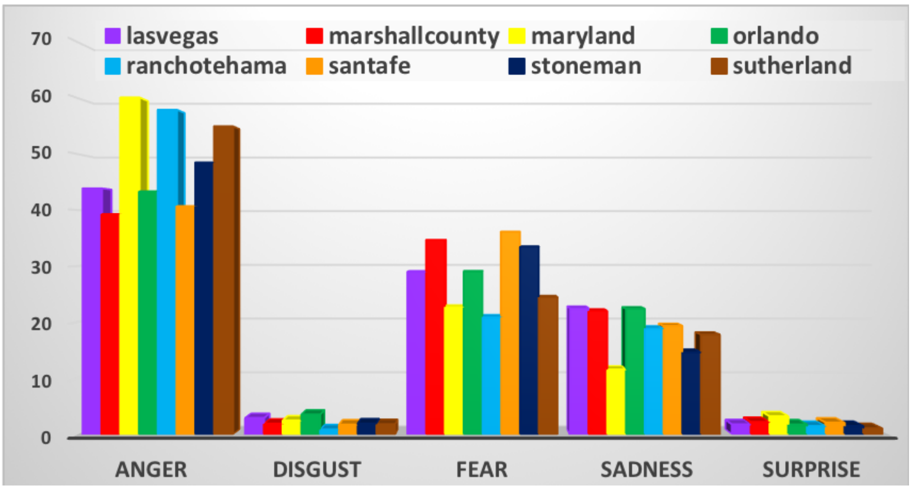


Figure 6.4: Emotion distribution for mass shooting events.
Source: The author.

sadness. The only exception is the event in Manchester, where the predominant emotion was fear. We also noticed that the New York December event (NY-december) evoked less sadness and more anger, compared to the other events. A possible explanation is that this attack has not resulted in any deaths.

A similar analysis on mass shooting events corroborates these findings. As shown in Figure 6.4, anger is the most evoked emotion, followed by fear and sadness. This pattern occurs for all events with no exception. We observed that, comparatively, less sadness and more anger are evoked for the event in Maryland. Like the NY-december event mentioned above, this event was the one with less victims (sum of deaths and injuries) among all.

We used demographics to help understand our findings, using only the three prevalent emotions. For terrorism events, Figure 6.6 details gender distribution, and Figure 6.5 stratifies the distribution of emotions by gender. The Y axis represents, for each class, the percentage

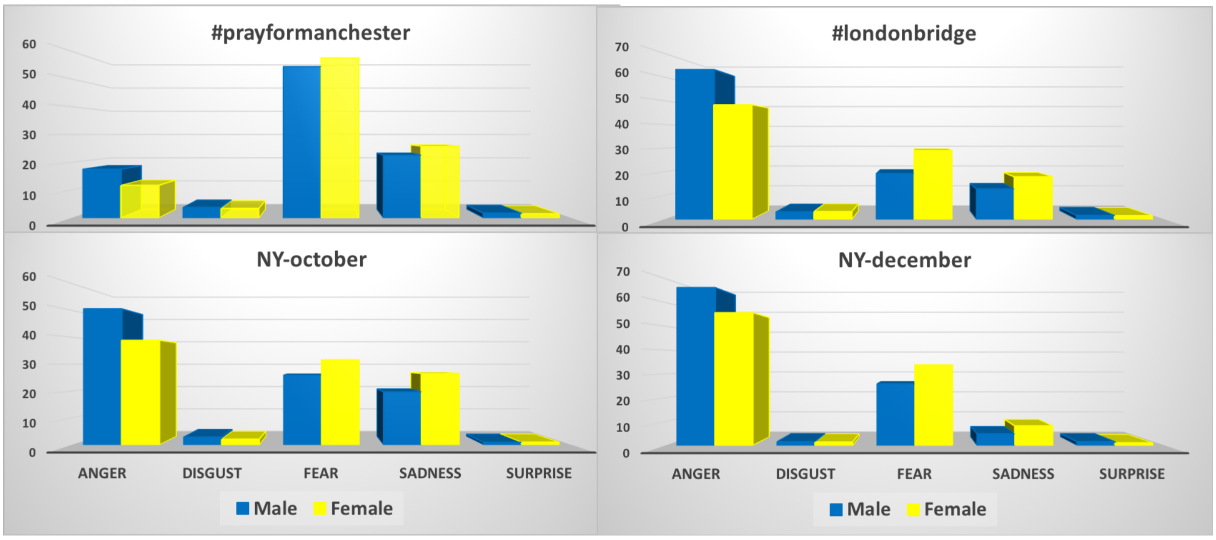


Figure 6.5: Emotion distribution by Gender for all terrorism events.
Source: The author.

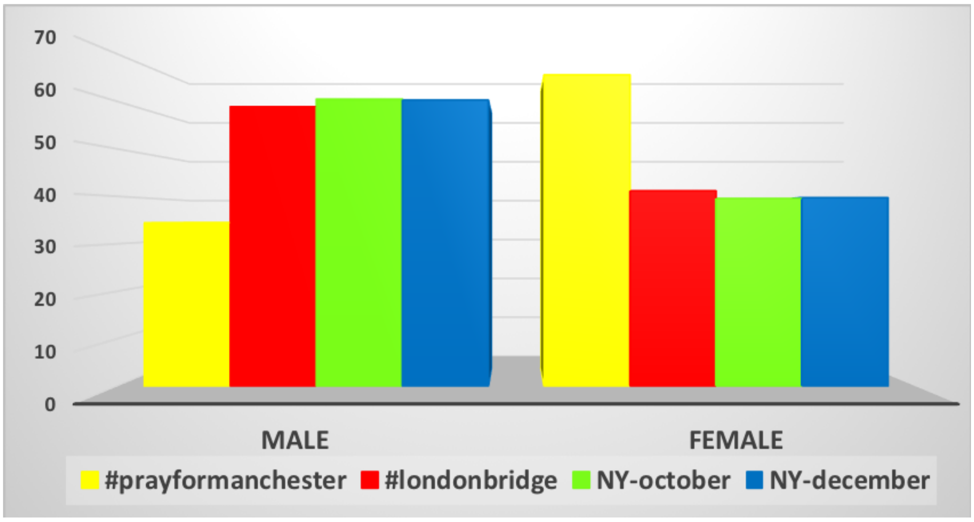


Figure 6.6: Tweet distribution by Gender for terrorism events.

with regard to the total number of tweets with emotion. Tweets in which user gender could not be determined were not considered. Our analysis on terrorism events showed that fear and sadness are more related to female users, whereas anger is proportionally more related to male users. We believe gender demographics partially explain the differences of the emotion distribution per event. In the Manchester event, tweeters were mostly women, possibly because the singer Ariana Grande is very popular in this demographic. In the events of London and New York, we observed that the tweeters were mostly male, the gender in which measures of anger are higher. A possible explanation is that the latter three events have affected the average citizen, who could potentially be at the location of the attack.

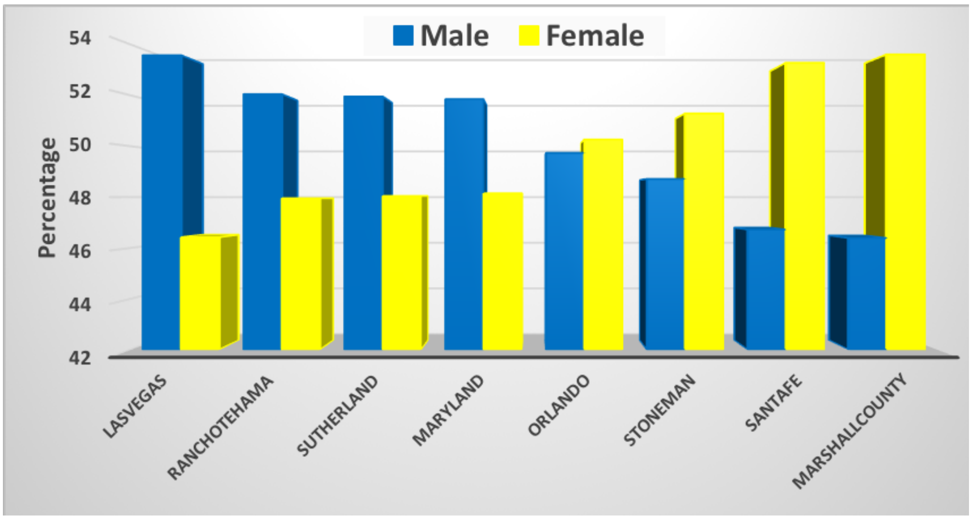


Figure 6.7: Tweet distribution by Gender for mass shooting events.
Source: The author.

The same emotion pattern with regard to gender was observed for the mass shooting events. For each event, Figure 6.7 displays the gender distribution and Figure 6.8, the gender distribution per emotion. We confirmed that fear and sadness are more related to women, and anger to men. By analyzing gender distribution in Figure 6.8, we found that engagement by Female users is considerably higher in three mass shooting events: Marshall County, Santa Fe and Stoneman. These three events occurred in schools, and a possible explanation may be that women relate more to these events due to the stronger ties between mothers and their children, yielding the necessity of venting their sentiments and sharing thoughts with the community. Rancho Tehama is an event that started in a school and spread to the whole community as it progressed, and this may explain a different behavior.

We also analyzed age distribution per event, and their relation to specific emotions. Figures 6.9 and 6.10 show the engagement by age groups for terrorism and mass shooting events, respectively. In both contexts, the engagement is higher in the age group between 35 and 54 years old. For terrorism events, this age range accounts for 50-60% of total engagement in three events. The only exception is Manchester, in which younger groups have a notable superior engagement. Even so, the 35-54 age group is still the most engaged one (about 30%). The engagement according to age groups is more smoothly distributed for mass shooting events (Figure 6.10). Although the 35-54 group is also the most active one (in average, 30%), its difference with regard to younger ones is much smaller. This pattern is observed across all events. The only slightly different behavior is observed for Santa Fe, in which the age groups 19-25 and 35-54 have almost the same level of engagement.

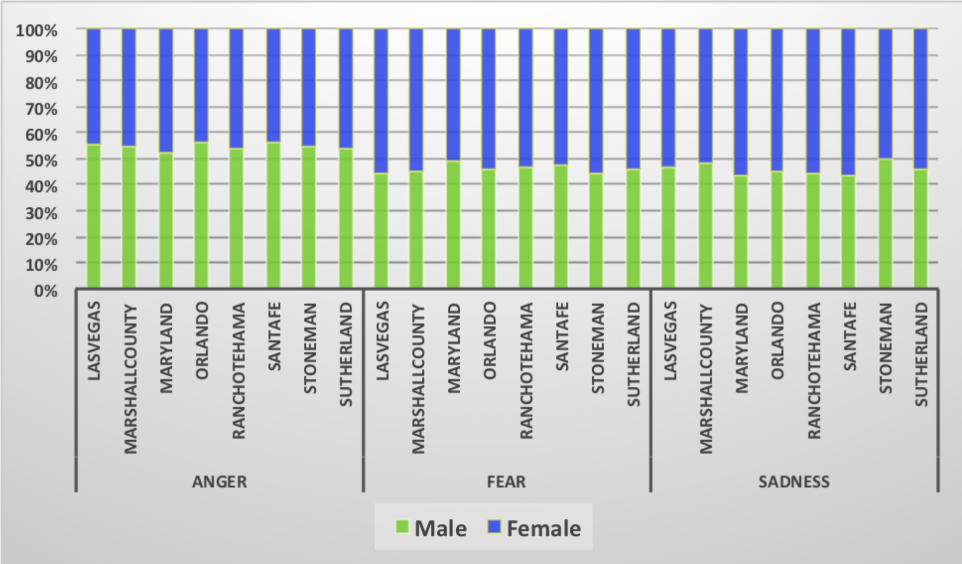


Figure 6.8: Gender distribution by emotion for mass shooting events.
Source: The author.

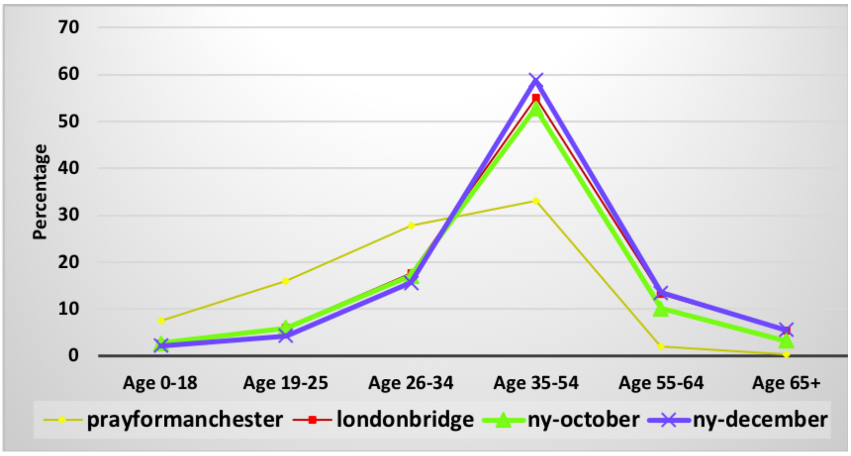


Figure 6.9: Engagement by age for terrorism events.
Source: The author.

For each event, we also compared age distribution by emotion. Figure 6.12 illustrates such comparison for terrorism events, where the Y axis represents, for each class, the percentage with regard to the total number of tweets with emotion. It is possible to observe a pattern concerning the UK events. As the age increases, the feeling of anger increases proportionally. Fear, on the other hand, is higher for young ages, and it smoothly drops as age increases. Regarding the US events, starting from age 19, there is a tendency of anger to remain constant, while fear has the tendency to drop as age increases. Sadness showed no clear behavior in any event. By analyzing the age distributions per emotion, we observed that the median age related to fear (36) is smaller compared to median age for anger (42), and so is the third quartile (3Q) values for age (45 compared to 51). However, when analyzed per event, we could not

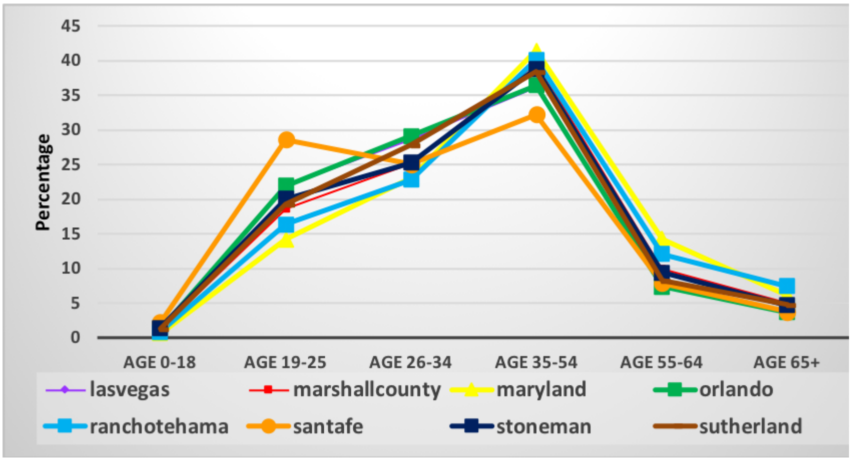


Figure 6.10: Engagement by age for mass shooting events.
Source: The author.

observe a pattern across the events and thus the influence of age on the emotional reaction was inconclusive.

Regarding the mass shooting events, the overall median age for Anger is 34, whereas for Fear it is 32, and the 3Q values are 46 and 43, respectively. Figure 6.13 shows age distribution by emotion, for each event. By comparing median and 3Q values related to Anger and Fear per event, we can observe that the users related to the latter are slightly younger. The only exceptions are for Rancho Tehama and Sutherland, where the 3Q values are comparable. No specific pattern was found for sadness. By plotting graphs showing the engagement per age group by emotion, shown in Figure 6.11, we observed for anger the same pattern identified for the UK events, i.e. as the age increases, the feeling of anger increases proportionally. Also, with two exceptions, the highest levels of engagement for fear are observed in people under 25, and then it decreases. The exceptions are Rancho Tehama and Marshall County, where this peak is observed at the range 26-34. Even so, these patterns involving age are more subtle, compared to gender, leading to the conclusion that it has less influence on the expressed emotions in such a context.

With all of our analysis, we conclude that each mass violent event may raise distinct predominant emotions according to the demographics, although anger is the prevailing emotion. In the analyzed events, gender was definitely more influential on the emotion than age. Fear and sadness are more related to the female gender, and anger, to the male one. Comparatively, fear is more related to younger populations, whereas the population that expressed anger is slightly older. As age progresses, we observed that the population that evokes anger increases slightly, and conversely, the population that evokes anger decreases. However, these patterns are quite subtle, and deserve further investigation.

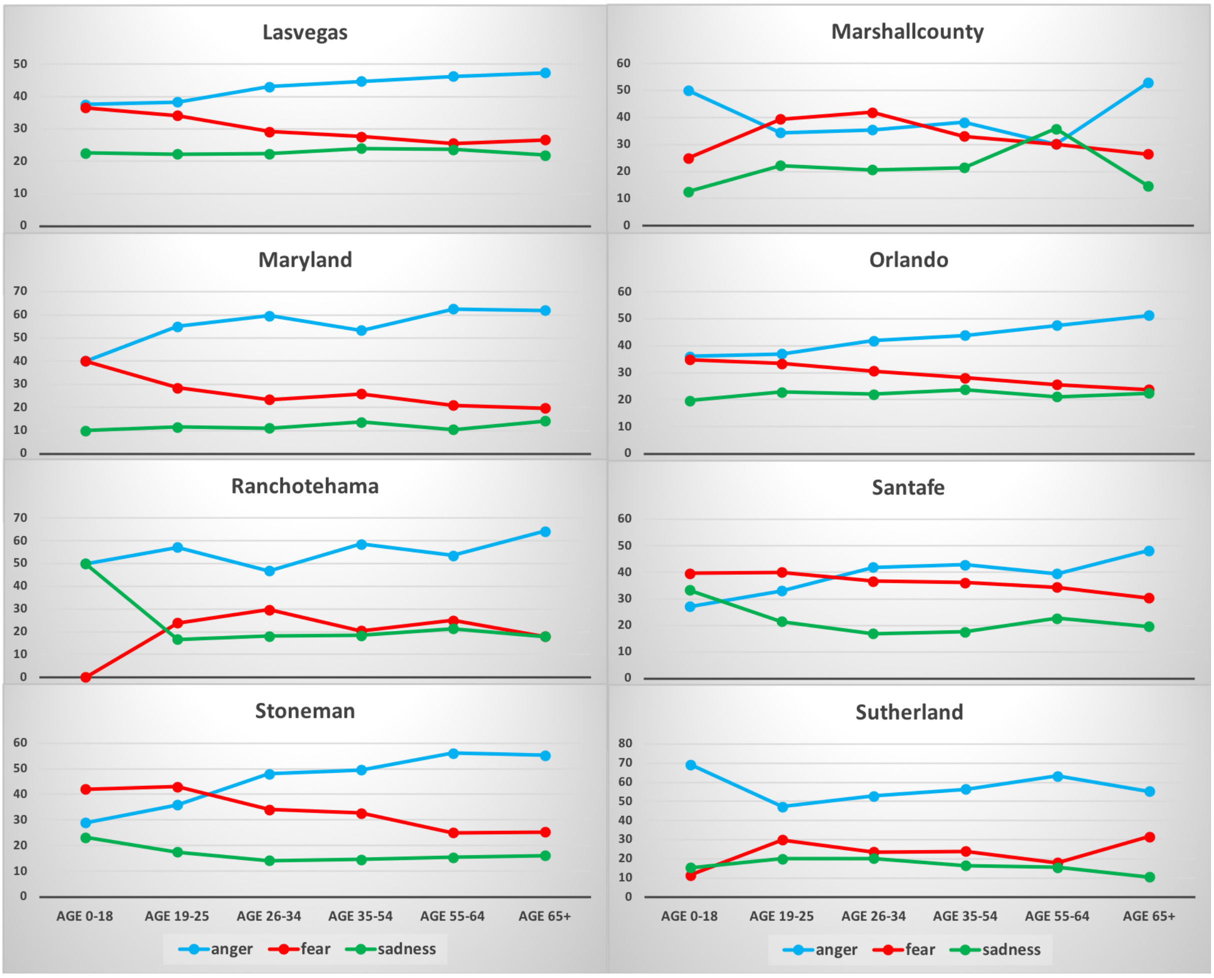


Figure 6.11: Engagement per age group by emotion for each mass shooting event.
Source: The author.

6.3 Q3: Does the proximity to the event influence the emotional reaction?

To answer this question, we compared the distribution of post-event tweets by emotion for each defined location category (see Section 4.2).

For terrorism, we compared UK against US tweets, and New York tweets against tweets related to other locations in US. We disregarded the category "other locations" due to the small number of tweets. These comparisons are displayed in Figure 6.14, and our results revealed no noticeable difference.

With regard to the mass shooting events, the scenario changed. In such a context, we were able to consider a thinner granularity for location, comparing tweets from the locations where the events took place (referred to as "local") against tweets within the US, as well as

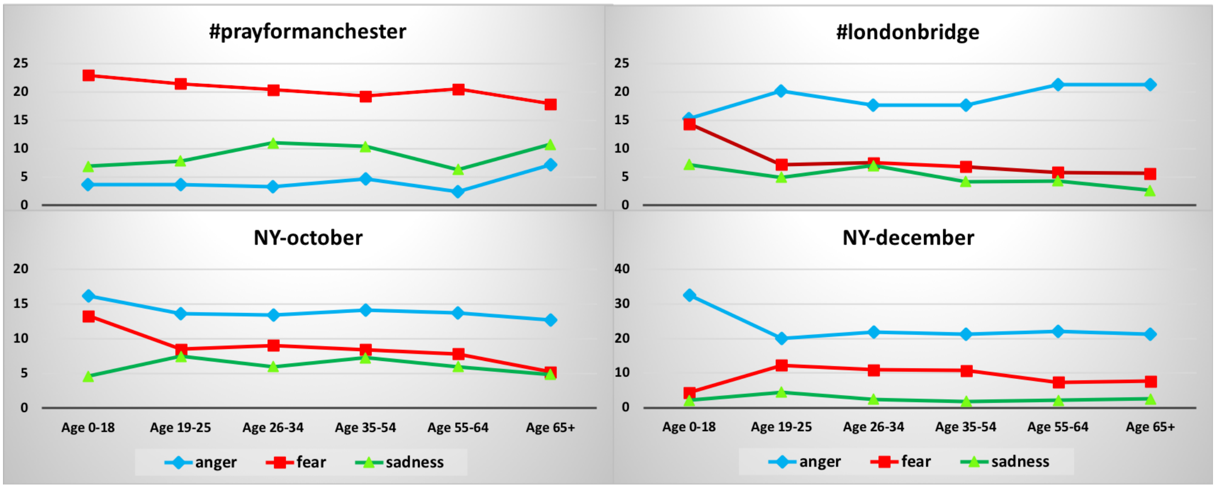


Figure 6.12: Engagement per age group by emotion for each terrorism event.
Source: The author.

outside the US ("other locations"). These comparisons are displayed in Figure 6.15. It can be noted that the prevalence of emotions in *local* tweets are distinct compared to the ones in the other two categories. For local tweets, measures of anger are comparatively lower, and the ones for sadness, higher. An exception is Las Vegas, in which sadness can be considered comparable in all locations. Fear has also comparatively higher measures in the local category in five events, or comparable levels otherwise, particularly with regard to locations outside the US. We consider that, in most cases, US and non-US tweets evoke comparatively the same emotions in most cases.

These findings are evidence that users closer to the locations where the mass shooting events took place tend to express more sadness and fear as they may be more related to the victims and the community, or even be victims themselves. On the other hand, users outside these communities tend to express more anger, as a response to the repercussion publicized on news or social media, and possibly because the events did not directly impact their local communities.

We conclude therefore that the location does influence the emotion expressed in mass shooting events, but no strong evidence was found for terrorism. Nevertheless, we believe that terrorism events may have shown a similar pattern (as indicated in the NY-october event) if we were able to apply the same granularity levels to group the locations.

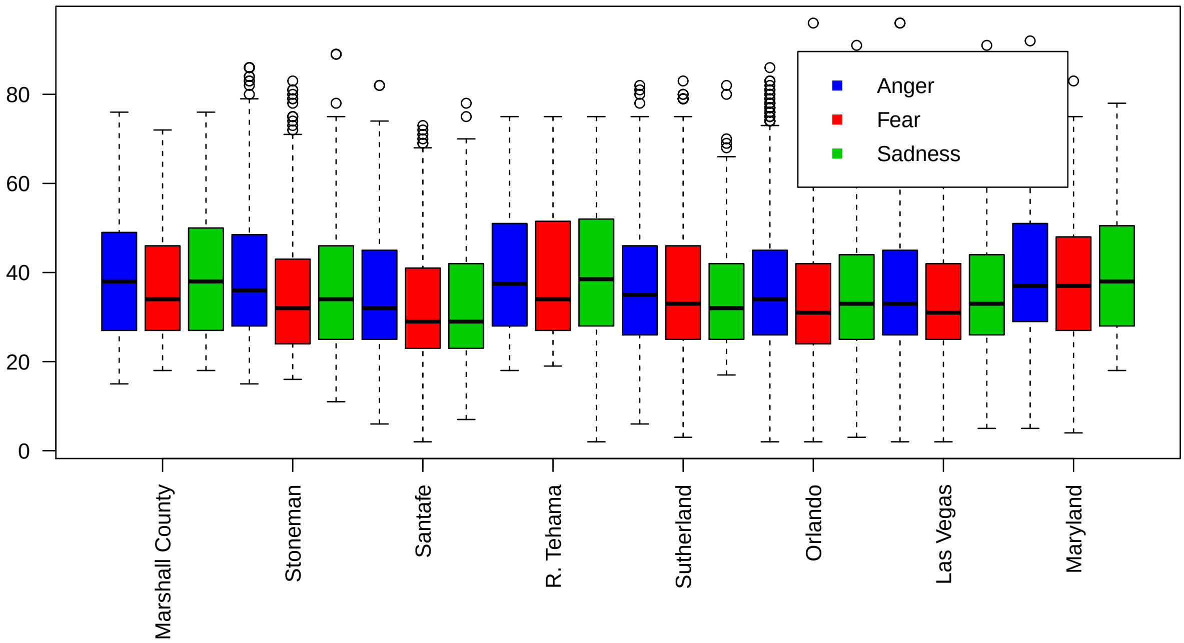


Figure 6.13: Age distribution by emotion for each mass shooting event.
Source: The author.

6.4 Q4: Does the number of people affected by the event have an impact on the emotional reaction?

To answer this question, we analyzed whether or not the number and type of victims affect the type of emotions expressed using post-event tweets. Figure 6.16 displays, for terrorism events, the amount of anger, sadness, and fear per event. The Y axis represents, for each class (number of deaths/injuries), the percentage of its total number of tweets distributed into emotions. Each class of deaths/injuries represents the number and type of victims of each event, and events displayed in the X axis are ordered by number of victims.

Our analysis of terrorism events showed that the level of sadness seems to be directly related to the number of victims (sum of deaths and injuries), whereas anger is inversely related. Fear is highly present when the number of victims is considerably high, but remains constant otherwise. We raised two hypotheses for this behavior. One hypothesis was that sadness may be a feeling more related to an event that victimizes more people, possibly because people feel helpless, or due to a sense of solidarity toward their families and close ones. Conversely, anger may be more present when tweeting about events that have fewer victims, due to the potential of harm. A second hypothesis is that these results are biased by the user's gender, as in Manchester,

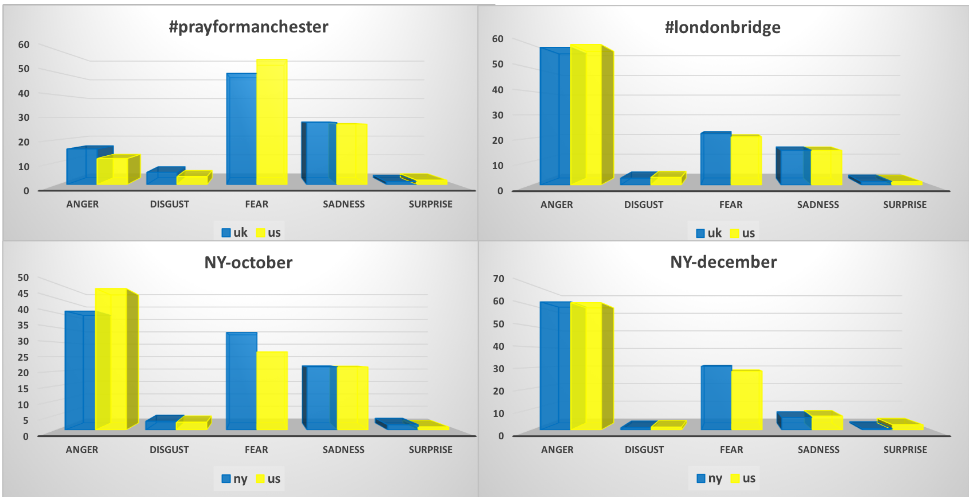


Figure 6.14: Emotion distribution by location for terrorism events.
Source: The author.

which has the highest number of casualties but is also related to a significant number of female users.

The first hypothesis was not confirmed by the analysis of the mass shooting events. As it can be seen in Figure 6.17, no clear pattern can be drawn between the number/type of victims and the expressed emotions. However, we could confirm the gender bias over number of casualties, considering the proportion of anger/fear expressed in some specific events; namely Santa Fe, Marshall County and Stoneman, which are the ones with the highest female engagement.

We observed in both contexts that the events with the least victims (NY-December and Maryland) are associated with the lowest level of sadness, and the highest level of anger. In the terrorism case, there are smooth decreasing/increasing trends for these emotions, respectively. However, these trends are not conclusive in the mass shooting context, due to the influence of gender.

In conclusion, we did not find strong evidence that the number of victims influences the emotion expressed, and we believe that the behaviors found are biased by the gender engagement. This issue deserves further investigation.

6.5 Q5: Are there differences in the expression of emotions depending on the event?

To answer this question, we searched for common words used in tweets related to the three prevailing emotions by exploring word clouds of the most frequent words in each category.

Figure 6.18 depicts the word clouds by emotion for the terrorism events using the top-50 frequent words. We found patterns related to the events themselves (e.g. location and nature

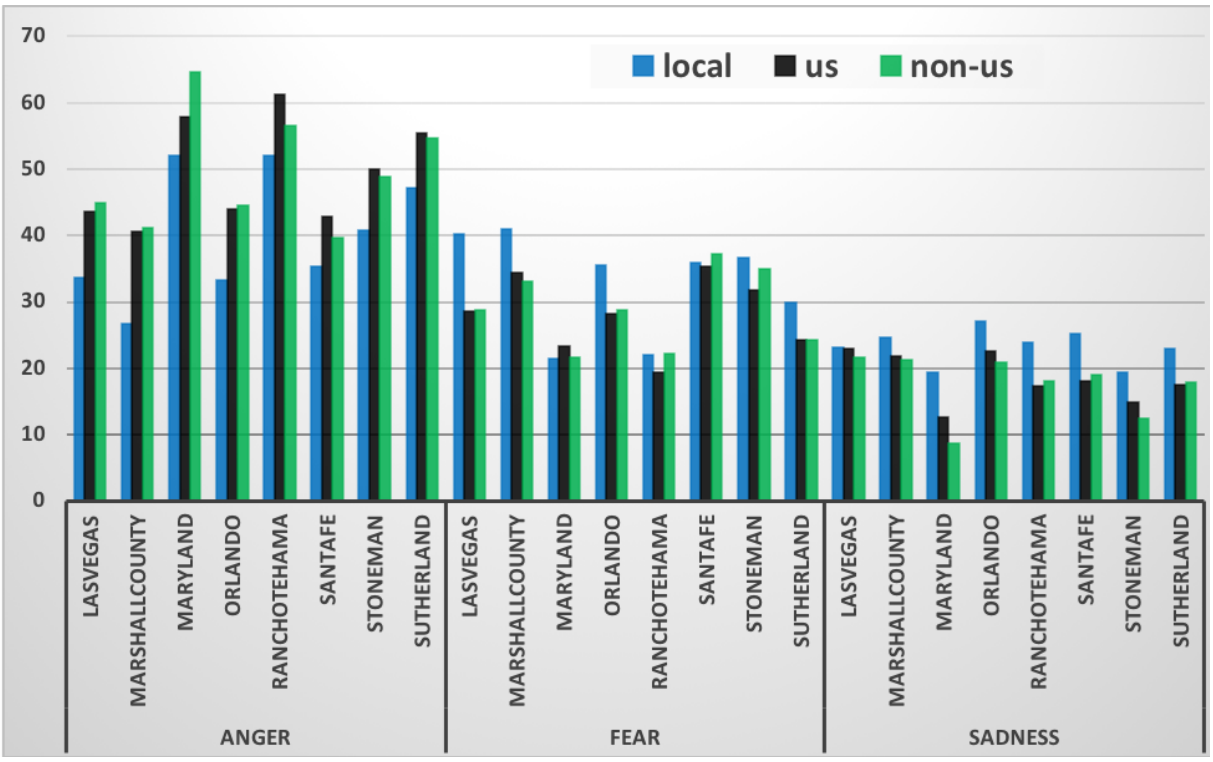


Figure 6.15: Emotion distribution by location for mass shooting events.
Source: The author.

of the event), emotions and generic words referring to mass violence in the form of terrorism. We noted that many of the keywords defined for filtering training seeds in our model were present in the respective emotion word clouds (e.g. "fuck" for anger, "worried" and "scared" for fear, "sad" and "saddened" for sadness). This is an evidence that the used keywords are indeed representative.

We also found common terms across all clouds that relate to terrorism events in general (e.g. "attack") and words related to religion (e.g. "muslim", "muslims", "islamic", "islam"), regardless of the emotion. Indeed, many people believe that Muslim extremists are behind these attacks, either as declared practitioners or as recruited members of an extremist religious terrorist group. There are also neutral terms describing the respective event location, such as "manchester" and "concert", "london" and "bridge", "new" and "york". The nature of the attack is also described accordingly, using "bomb", "bomber", "explosion", and "van".

Concerning the clouds for fear and sadness, we observe more frequent words that demonstrate solidarity such as "pray", "prayers" and "condolences". We suppose that tweeters expressing fear and sadness are likely to relate to the victims and their close ones, putting themselves in their shoes and sharing words of affection and support. On the other hand, clouds for anger show with frequency words of hate and intolerance, such as "stop", "terror", and "awful".

For the US terrorism events, the word "trump" (referencing the US president) was

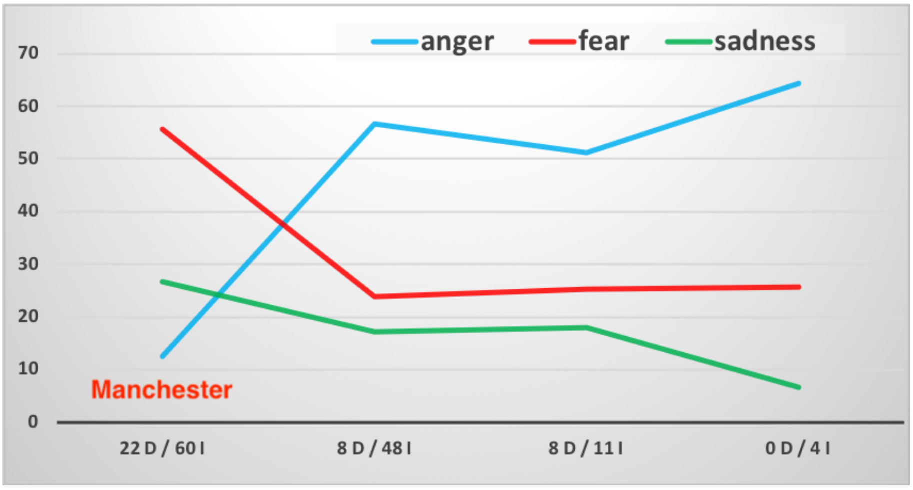


Figure 6.16: Emotion by number of victims - deaths(D)/injuries(I) - in terrorism events. Source: The author.

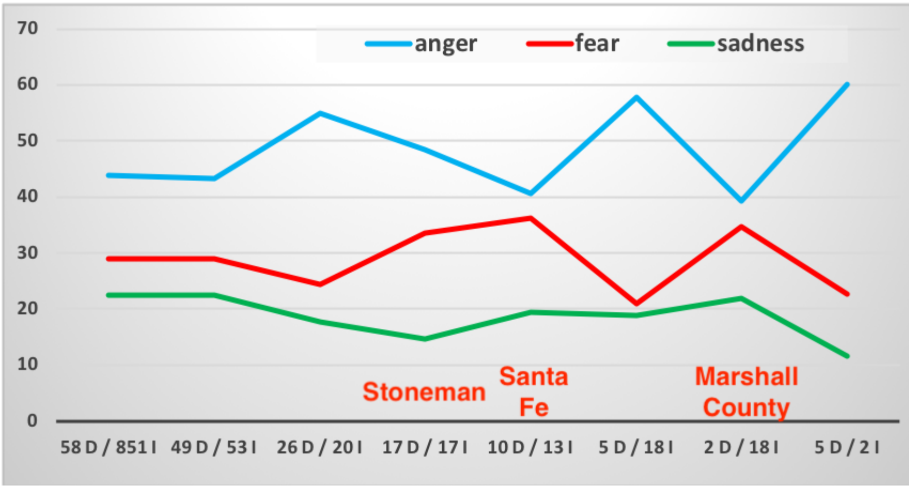
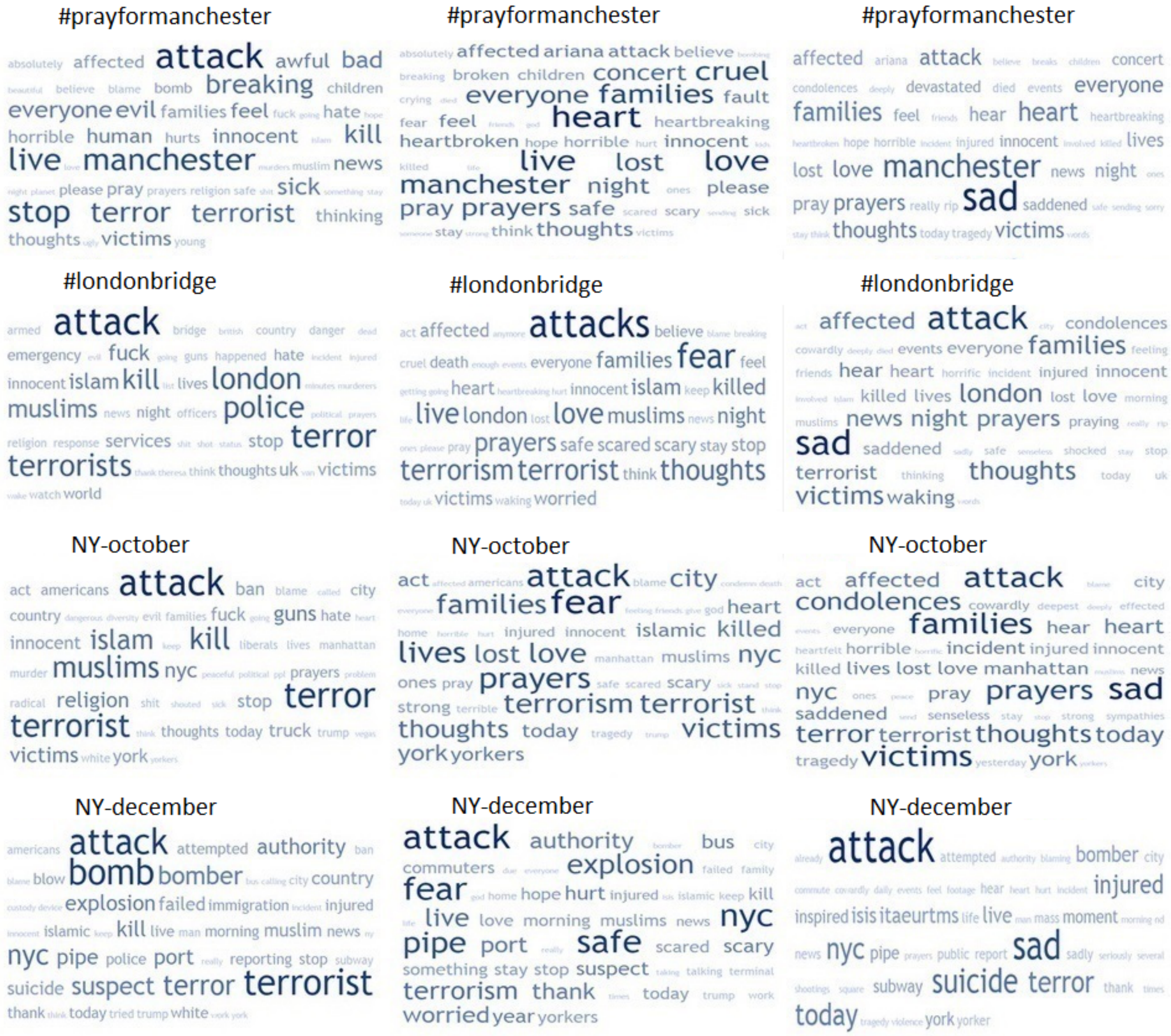


Figure 6.17: Emotion by number of victims - deaths(D)/injuries(I) - in mass shooting events. Source: The author.

present in 4 out of the 6 word clouds and "obama" (former US president) appeared once. Such a political mention appears just once for UK events ("political"). This might show a cultural difference between US and UK tweeters, where Americans are more engaged in claiming safety to their President (or blaming him).

We did the same for mass shooting events, examining the top-50 most frequent words, which are displayed in Appendix A due to their size. Figure 6.19 depicts the word clouds using only the top-20 frequent words. To display more representative words, we generated these clouds excluding any reference to locations (e.g. "kentucky", "vegas").

Similar conclusions can be drawn from mass shooting word clouds, in comparison to the terrorism events ones. First, we confirmed the presence of the keywords used to filter the training seeds (e.g. "fuck", "fucking", "sad", "fear"). As can be noted in the clouds depicted in Appendix A, all events, regardless the emotion, contained references to the respective location



Anger Fear Sadness

Figure 6.18: Word Clouds for all terrorism events, by emotion.
Source: The author.

in different granularities (e.g. cities such as "vegas", "orlando" or states such as "california", "texas"). Words describing the type of place are also present (e.g. "school", "church").

All clouds contained expressions related to mass shootings (e.g. "shooting", "mass", "schoolshooting"). Targets of the events are also described, such as "kids" and "children" for school shootings, "journalists" for the press agency, "gays" for the nightclub event, and generic words, such as "victims" and "people", to events that also targeted random population (e.g. Las Vegas).

Concerning the expression of emotions, similarly to terrorism events, words of solidarity and empathy to the victims and their loved ones are common to clouds of fear and sadness, such as "families", "prayers", "heart", "thoughts", and "love". For fear, we also find words of wondering and questioning (e.g. "happening", "happened"). Words of hate and negativity are

Figure 6.19: Word Clouds for all mass shooting events
Source: The author.

linked to clouds of anger, such as "violence", "stop", "mental", "white", and even references to terrorism (e.g. "terrorism", "terror").

Considering the top-50 frequent words, the word "gun" appears in all clouds, but with a different weight. This is clear in Figure 6.19, that displays only the top-20, where "gun" is basically related to anger only, and often co-occurs with "law". A possible explanation is that anger evokes a pro/against gun stance debate, as people might relate such tragedies to the legally armed population, or to laws that restrict the population to possess firepower, which prohibits fast responsive actions if not through police force. This hypothesis is reinforced by references to the US president in three events (Maryland, Orlando and Sutherland) for anger clouds. This is also evidence, as for the terrorism events in NY, that the communities in the US are likely to relate political people to such tragedies, possibly by blaming them or claiming for safety.

We conclude that, respected the differences of each context, the way users express their sentiments for each emotion is similar, showing that mass violent events in general are likely to raise similar reactions from the population. Whereas fear and sadness seem to be more related to the consequences of the attacks, evoking solidarity and empathy, anger appears to be more associated to their causes, using strong words of hate, intolerance and call for justice.

7 CONCLUSION AND FUTURE WORKS

This work described a study on the emotional behavior of Twitter users in reaction to mass violent events. We covered four terrorism events and eight mass shooting incidents, thus not restricting ourselves to a specific context. To understand emotion shifts, we compared pre-event tweets related to the respective community, and tweets referring to the event. Sentiment analysis encompassed a set of negative emotions, rather than simply sentiment polarity, and we applied deep learning for emotion prediction by training a CNN as an alternative to traditional machine learning approaches and feature engineering. To better understand our results, we explored demographic data such as location, age, and gender, extracted from users profile with use of auxiliary tools.

To overcome the lack of large, quality annotated sets in these domains, we trained the CNN using seeds that were automatically selected. We experimented on different strategies for automatically selecting the training seeds, and we chose the strategy of filtering by keywords, which yielded better performance. We developed experiments to verify which emotion prediction model would perform better for each kind of event. Models varied on the source of the training seeds, and evaluation was conducted against our gold standards, one of the contributions of our work.

In this work we analyzed the sentiment using both pre/post event tweets from the general population, thus not restricting ourselves to a specific sampled group. In addition, we used information extracted from users' profiles to analyze emotions according to users' demographics, more specifically location, gender, and age, a novel aspect of sentiment analysis in this domain. Our study concluded that the emotional reactions to both terrorism and mass shooting events present similar patterns, thus showing how they potentially impact people within and outside the affected community in a similar way. Our results showed that when a mass violent events occurs, a shift of emotion towards anger, fear and sadness can be noticed. In addition, our demographic analysis has shown that gender has influence on tweeters reactions. Our data indicated that measures for fear and sadness are proportionally higher for Women, whereas anger is proportionally more related to the Male demographics. Age is less influential, but our analysis has shown that as the age increases, the sentiment of anger increases proportionally, and that fear is related to slightly younger populations. The influence of the proximity to the event on the expressed emotions could only be observed for mass shooting events. In these events, users closer to the location of the events tend to express more sadness and fear if compared to users from outside the targeted communities, where anger is higher instead. The word clouds we created showed that words identifying event location, nature of the attack and emotions were commonly used. Emotions such as fear and sadness evoke the use of words that demonstrate solidarity, and emotions such as anger evoke the use of words of hate, intolerance and call for justice. Mentions to politicians were common as also references to gun control law in mass shooting incidents and religion in terrorism events, showing how the communities are likely to link debatable subjects to such tragedies. Lastly, our analysis on the influence of number and type of casualties was inconclusive.

Limitations of our work include the limited size of our gold standards, which may influence on the experiments evaluation results. The use of auxiliary tools also deserves further investigation on other alternatives and approaches. Our tweet sampling method is simple and could be more stressed in order to be able to yield better results. Finally, some of our analysis may have a bias due to the demographics, and unbalanced number of tweets with regards to pre/post event classes, and emotion classes.

Our study has been published twice (HARB; BECKER, 2018; HARB; BECKER, 2019), it has been awarded as the best paper in one conference (HARB; BECKER, 2018), and has been invited to have an extended version submitted to the special issues of the Information Systems journal (HARB; BECKER, 2019). Our research confirmed the value of the information Twitter users provide, and has also shown how such information can be exploited in order to understand relevant phenomena such as mass violent events. It constitutes a first step towards understanding the emotional reactions such events evoke on population, and we hope it encourages further studies on the subject. This type of result might be used in developing specific assistance programs for coping with the psychological effects of such tragedies, and to help the population to cope with the constant sense of threat and fear effectively. Future work includes: a) the extension of the CNN to address tweets in a multilingual environment, thus reaching a larger number of events, countries and population; b) to improve the techniques to extract the demographics from users profiles; c) to employ topic analysis on tweets to further investigate the contents exchanged; among others.

REFERENCES

AGARWAL, A. et al. Sentiment analysis of twitter data. In: **Proceedings of the Workshop on Languages in Social Media**. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. (LSM '11), p. 30–38. ISBN 978-1-932432-96-1. Available from Internet: <http://dl.acm.org/citation.cfm?id=2021109.2021114>.

AGGRAWAL, A. Mass murder. In: PAYNE-JAMES, J. J. et al. (Ed.). **Encyclopedia of Forensic and Legal Medicine**. [S.l.]: Elsevier Academic Press, 2005. p. 216–223.

AIN, Q. T. et al. Sentiment analysis using deep learning techniques: A review. **International Journal of Advanced Computer Science and Applications**, The Science and Information Organization, v. 8, n. 6, 2017.

AZIZAN, S. A.; AZIZ, I. A. Terrorism Detection Based on Sentiment Analysis Using Machine Learning. **Journal of Engineering and Applied Sciences**, v. 12, n. 3, p. 691–698, 2017.

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: **Proc. of the Intl. Conf. on Language Resources and Evaluation (LREC)**. [S.l.: s.n.], 2010. v. 10, n. 2010, p. 2200–2204.

BURNAP, P. et al. Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. **Social Network Analysis and Mining**, v. 4, n. 1, p. 206, Jun 2014.

CHONG, M. Sentiment analysis and topic extraction of the twitter network of #prayforparis. In: **Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives Through Information & Technology**. Silver Springs, MD, USA: American Society for Information Science, 2016. (ASIST '16), p. 133:1–133:4. Available from Internet: <http://dl.acm.org/citation.cfm?id=3017447.3017580>.

CHOUDHURY, M. D. et al. Social media participation in an activist movement for racial equality. In: **Proc. of the 10th Intl. Conference on Web and Social Media (ICWSM)**. [S.l.: s.n.], 2016. p. 92–101.

COHEN-LOUCK, K.; BEN-DAVID, S. Coping with terrorism: Coping types and effectiveness. **International Journal of Stress Management**, v. 24, n. 1, p. 1–17, 2017.

COLLOBERT, R. et al. Natural language processing (almost) from scratch. **Journal of Machine Learning Research**, v. 12, p. 2493–2537, 2011.

CRENSHAW, M. The causes of terrorism. **Comparative Politics**, Comparative Politics, Ph.D. Programs in Political Science, City University of New York, v. 13, n. 4, p. 379–399, 1981. ISSN 00104159.

CREPEAU-HOBSON, F. et al. A coordinated mental health crisis response: Lessons learned from three colorado school shootings. **Journal of School Violence**, v. 11, n. 3, p. 207–225, 2012.

DESAI, M.; MEHTA, M. A. Techniques for sentiment analysis of twitter data: A comprehensive survey. In: **2016 International Conference on Computing, Communication and Automation (ICCCA)**. [S.l.: s.n.], 2016. p. 149–154.

EKMAN, P.; FRIESEN, W. Emotion in the human face system. **Cambridge University Press, San Francisco, CA,**, 1982.

ELSHERIEF, M.; BELDING, E. M.; NGUYEN, D. # notokay: Understanding gender-based violence in social media. In: **Proc. of the 11th Intl. Conference on Web and Social Media (ICWSM)**. [S.l.: s.n.], 2017. p. 52–61.

EMOTION: Theory, Research and Experience. Volume 1. Theories of Emotion. Edited by R. Plutchik and H. Kellerman. (Pp. 399; illustrated; £19.00.) Academic Press: London. 1980. **Psychological Medicine**, Cambridge University Press, v. 11, n. 1, p. 207–207, 1981.

FAN, H. et al. Learning deep face representation. **CoRR**, 2014. Available from Internet: <http://arxiv.org/abs/1403.2802>.

GALLEGOS, L. et al. Geography of emotion: Where in a city are people happier? In: **Proc. of the 25th Intl. Conf. on World Wide Web (WWW)**. [S.l.: s.n.], 2016. p. 569–574.

GARG, P.; GARG, H.; RANGA, V. Sentiment analysis of the Uri terror attack using twitter. In: **Proc. of the Intl. Conf. on Computing, Communication and Automation (ICCCA)**. [S.l.: s.n.], 2017. p. 17–20.

GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. **Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group**, p. 1–6, 2009. Available from Internet: <http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf>.

GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. p. 1–6, 2009. Available from Internet: <http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

HARB, J. G. D.; BECKER, K. Emotion analysis of reaction to terrorism on twitter. In: **Proc. of the SBC Brazilian Symposium on Databases**. [S.l.: s.n.], 2018. p. 97–108.

HARB, J. G. D.; BECKER, K. Comparing emotional reactions to terrorism events on twitter. In: OLIVEIRA, J. et al. (Ed.). **Big Social Data and Urban Computing**. [S.l.]: Springer, 2019. v. 926, chp. 7. *(To appear.)*.

HOFFMAN, B. **Inside terrorism**. 3rd ed.. ed. [S.l.]: Columbia University Press, 2013. ISBN 0231126980 0231126999 0231510462 9780231126991.

HOSSIN, M.; M.N, S. A review on evaluation metrics for data classification evaluations. **International Journal of Data Mining Knowledge Management Process**, v. 5, p. 01–11, 03 2015.

JONES, N. M. et al. Tweeting negative emotion: An investigation of twitter data in the aftermath of violence on college campuses. **Psychological Methods**, v. 21, p. 526–541, 2016.

KELLEHER, J.; NAMEE, B.; D'ARCY, A. **Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies**. MIT Press, 2015. (The MIT Press). ISBN 9780262029445. Available from Internet: <https://books.google.com.br/books?id=uZxOCgAAQBAJ>.

KHARDE, V. A.; SONAWANE, S. Article: Sentiment analysis of twitter data: A survey of techniques. **International Journal of Computer Applications**, v. 139, n. 11, p. 5–15, April 2016. Published by Foundation of Computer Science (FCS), NY, USA.

KIM, Y. Convolutional neural networks for sentence classification. In: **Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) - ACL**. [S.l.: s.n.], 2014. p. 1746–1751.

KOTSIANTIS, S. B. Supervised machine learning: A review of classification techniques. In: **Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies**. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2007. p. 3–24. ISBN 978-1-58603-780-2. Available from Internet: <http://dl.acm.org/citation.cfm?id=1566770.1566773>.

LEBRET, R.; COLLOBERT, R. Word embeddings through hellinger pca. In: **Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics**. Association for Computational Linguistics, 2014. p. 482–490. Available from Internet: <http://aclweb.org/anthology/E14-1051>.

LERMAN, K. et al. Emotions, demographics and sociability in twitter interactions. In: **Proc. of the 10th Intl. Conference on Web and Social Media (ICWSM)**. [S.l.: s.n.], 2016. p. 201–210.

LI, C. et al. Recursive deep learning for sentiment analysis over social data. In: **2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)**. [S.l.: s.n.], 2014. v. 2, p. 180–185.

LI, Q. et al. Real-time novel event detection from social media. In: **Proc. of the 33rd IEEE International Conference on Data Engineering (ICDE)**. [S.l.: s.n.], 2017. p. 1129–1139.

LIU, B. Sentiment analysis and opinion mining. **Synthesis Lectures on Human Language Technologies**, v. 5, n. 1, p. 1–167, 2012.

LIU, B.; HU, M.; CHENG, J. Opinion observer: Analyzing and comparing opinions on the web. In: **Proc. of the 14th Intl. Conf. on World Wide Web (WWW)**. [S.l.: s.n.], 2005. p. 342–351.

LOWE, S.; GALEA, S. The mental health consequences of mass shootings. **Trauma, Violence & Abuse**, v. 18, n. 1, p. 62–82, 06 2015.

LOWE, S. R.; BLACHMAN-FORSHAY, J.; KOENEN, K. C. Trauma as a public health issue: Epidemiology of trauma and trauma-related disorders. In: SCHNYDER, U.; CLOITRE, M. (Ed.). **Evidence Based Treatments for Trauma-Related Psychological Disorders: A Practical Guide for Clinicians**. [S.l.]: Springer, 2015. p. 11–40.

MAGUEN, S.; PAPA, A.; LITZ, B. T. Coping with the threat of terrorism: A review. **Anxiety, Stress, & Coping**, Routledge, v. 21, n. 1, p. 15–35, 2008.

MANSOUR, S. Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. **Procedia Computer Science**, Elsevier B.V., v. 140, p. 95–103, 2018.

MIKOLOV, T. et al. Efficient estimation of word representations in vector space. **CoRR**, abs/1301.3781, 2013. Available from Internet: <http://arxiv.org/abs/1301.3781>.

MIKOLOV, T. et al. Distributed representations of words and phrases and their compositionality. In: **Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2**. USA: Curran Associates Inc., 2013. (NIPS'13), p. 3111–3119. Available from Internet: <http://dl.acm.org/citation.cfm?id=2999792.2999959>.

MIRANI, T. B.; SASI, S. Sentiment analysis of isis related tweets using absolute location. In: **Proc. of the 2016 International Conference on Computational Science and Computational Intelligence (CSCI)**. [S.l.: s.n.], 2016. p. 1140–1145.

MITCHELL, L. et al. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. **PLOS ONE**, Public Library of Science, v. 8, n. 5, p. 1–15, 05 2013.

MOHAMMAD, S. #Emotional Tweets. In: **Proc. of the 1rst Joint Conf. on Lexical and Computational Semantics**. [S.l.: s.n.], 2012. p. 246–255.

MOHAMMAD, S.; KIRITCHENKO, S.; ZHU, X. NRC-canada: Building the state-of-the-art in sentiment analysis of tweets. In: **Proc. of the 7th Intl. Workshop on Semantic Evaluation (SEMEVAL)**. [S.l.: s.n.], 2013. p. 321–327.

MOHAMMAD, S. M. 9 - sentiment analysis: Detecting valence, emotions, and other affectual states from text. In: MEISELMAN, H. L. (Ed.). **Emotion Measurement**. Woodhead Publishing, 2016. p. 201 – 237. ISBN 978-0-08-100508-8. Available from Internet: <http://www.sciencedirect.com/science/article/pii/B9780081005088000096>.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. **Computational Intelligence**, v. 29, n. 3, p. 436–465, 2013.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. v. 29, n. 3, p. 436–465, 2013.

MOHAMMAD, S. M. et al. Sentiment, emotion, purpose, and style in electoral tweets. v. 51, n. 4, p. 480–499, 2015.

MUNEZERO, M. D. et al. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. **IEEE Transactions on Affective Computing**, v. 5, n. 2, p. 101–111, 2014.

MURPHY, K. P. **Machine Learning: A Probabilistic Perspective**. [S.l.]: The MIT Press, 2012. ISBN 0262018020, 9780262018029.

ORTONY, A.; TURNER, T. J. What's basic about basic emotions? **Psychological Review**, v. 97, n. 3, p. 315, 1990.

PENNINGTON, J.; SOCHER, R.; MANNING, C. D. GloVe: Global Vectors for Word Representation. In: **Empirical Methods in Natural Language Processing (EMNLP)**. [S.l.: s.n.], 2014. p. 1532–1543. <http://www.aclweb.org/anthology/D14-1162>. Accessed in: 2017-03-21.

PURVER, M.; BATTERSBY, S. Experimenting with Distant Supervision for Emotion Classification. **Proc. of the 13th Conference of the European Chapter of the Association for Computational Linguistics**, p. 482–491, 2012.

ROSER, M.; NAGDY, M.; RITCHIE, H. **"Terrorism"**. 2018. OurWordInData.org. Available from Internet: <https://ourworldindata.org/terrorism>.

RUSSELL, J. A.; MEHRABIAN, A. Evidence for a three-factor theory of emotions. **Journal of Research in Personality**, Elsevier, v. 11, n. 3, p. 273–294, 1977.

SAKAKI, T.; OKAZAKI, M.; MATSUO, Y. Earthquake shakes twitter users: Real-time event detection by social sensors. In: **Proc. of the 19th Intl. Conf. on World Wide Web (WWW)**. [S.l.: s.n.], 2010. p. 851–860.

SCHMIDHUBER, J. Deep learning in neural networks. **Neural Netw.**, Elsevier Science Ltd., Oxford, UK, UK, v. 61, n. C, p. 85–117, jan. 2015. ISSN 0893-6080. Available from Internet: <http://dx.doi.org/10.1016/j.neunet.2014.09.003>.

SEVERYN, A.; MOSCHITTI, A. On the automatic learning of sentiment lexicons. In: **Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. Association for Computational Linguistics, 2015. p. 1397–1402. Available from Internet: <http://aclweb.org/anthology/N15-1159>.

SEVERYN, A.; MOSCHITTI, A. Twitter sentiment analysis with deep convolutional neural networks. In: **Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval**. New York, NY, USA: ACM, 2015. (SIGIR ’15), p. 959–962. ISBN 978-1-4503-3621-5. Available from Internet: <http://doi.acm.org/10.1145/2766462.2767830>.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning: From Theory to Algorithms**. New York, NY, USA: Cambridge University Press, 2014. ISBN 1107057132, 9781107057135.

SHEN, Y. et al. Learning semantic representations using convolutional neural networks for web search. In: **Proceedings of the 23rd International Conference on World Wide Web**. New York, NY, USA: ACM, 2014. (WWW ’14 Companion), p. 373–374. ISBN 978-1-4503-2745-9. Available from Internet: <http://doi.acm.org/10.1145/2567948.2577348>.

SIMON, T. et al. Twitter in the cross fire—the use of social media in the westgate mall terror attack in kenya. **PloS One**, v. 9, Aug 2014.

SINGH, S.; CHOUDHARY, S. S. Social Media Data Analysis: Twitter Sentimental Analysis Using R Language. **International Journal of Advances in Electronics and Computer Science**, v. 4, n. 11, p. 13–17, 2017.

SOCHER, R. et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**. Association for Computational Linguistics, 2013. p. 1631–1642. Available from Internet: <http://www.aclweb.org/anthology/D13-1170>.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Inf. Process. Manage.**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 45, n. 4, p. 427–437, jul. 2009. ISSN 0306-4573. Available from Internet: <http://dx.doi.org/10.1016/j.ipm.2009.03.002>.

SUTTLES, J.; IDE, N. Distant supervision for emotion classification with discrete binary values. In: GELBUKH, A. (Ed.). **Computational Linguistics and Intelligent Text Processing**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 121–136. ISBN 978-3-642-37256-8.

TABOADA, M. et al. Lexicon-based methods for sentiment analysis. **Computational Linguistics**, v. 37, n. 2, p. 267–307, 2011. Available from Internet: <https://doi.org/10.1162/COLI_a_00049>.

VARGAS, S. et al. Comparing overall and targeted sentiments in social media during crises. In: **Proc. of the 10th Intl. Conference on Web and Social Media (ICWSM)**. [S.l.: s.n.], 2016. p. 695–698.

WANG, N.; VARGHESE, B.; DONNELLY, P. D. A machine learning analysis of twitter sentiment to the sandy hook shootings. In: **Proc. of the IEEE 12th Intl. Conf. on e-Science (e-Science)**. [S.l.: s.n.], 2016. p. 303–312.

WANG, S.; YAO, X. Multiclass imbalance problems: Analysis and potential solutions. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, v. 42, n. 4, p. 1119–1130, Aug 2012. ISSN 1083-4419.

WANG, W. et al. Harnessing twitter 'big data' for automatic emotion identification. In: **Proc. of the 2012 ASE/IEEE International Conference on Social Computing**. [S.l.: s.n.], 2012. p. 587–592.

ZHANG, G. P. Neural networks for classification: a survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 30, n. 4, p. 451–462, Nov 2000. ISSN 1094-6977.

ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. **Wiley Interdiscip. Rev. Data Min. Knowl. Discov.**, v. 8, n. 4, 2018.

ZIMBRA, D. et al. The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. **ACM Trans. Management Inf. Syst.**, v. 9, n. 2, p. 5:1–5:29, 2018.

Appendices

AppendixA

In this Appendix, the top 50 word clouds for the mass shooting events with regard to the three prevailing emotions (anger, fear, and sadness) are displayed respectively in figures A.1, A.2, A.3.

Figure A.1: Top-50 anger word clouds for mass shooting events

Figure A.2: Top-50 fear word clouds for mass shooting events

Figure A.3: Top-50 sadness word clouds for mass shooting events