Research Proposal

#BlackLivesMatter: Decomposing the Social Movement and the Stories Behind

Yee Torng, **Hoo (P101447)**
University Kebangsaan Malaysia
Fakulti Teknologi dan Sains Maklumat
p101447@siswa.ukm.edu.my

1. Introduction

For decades, researchers have been trying to understand the social movements. However, it has not been an easy task to do so, where it involves many factors could interfere the social movement. Mohamed Bouazizi[1], a fruit monger burned himself in front of the town hall of EI Qued on 17 December 2010, due to unfair treatment, harassment and humiliation by some official. The death of Bouaziz triggered the anger that had been suppressed for long due to the poverty, monarchy, corruption etc. eventually Arab Spring Social Movement formed and spread across over Arab. One year later, London city was in crisis that shocked the world. Thousands of people rioted in cities and towns, looting and arson that caused 5 deaths. The London Riots was initiated by dead of Mark Dunggan[2] who was shot dead by police on 4th Aug 2011. *Fingers were pointed at increasing social inequalities, growing alienation among the young, and the poor example set by the greed of bankers and of MPs' fraudulent expenses* (***Lewis et. al., 2011***). March 2019, China government passed a bill regarding extradition to amend the fugitive offender's ordinance in Hong Kong. Two to three months later the social movements started to form against the bill and requested the Hong Kong Chief Executive Carrier Lim to step down and withdraw the complete bill as the protestors believe the bill erosion of Hong Kong's legal system and damage the business climate.

From the three social movements in the past mentioned above, each of the social movement started or triggered by some tragic event (Death of Bouaziz and Dunggan) or the political move or bill against the will of the people. Then people walk to the streets to

[1] *Tarek el-Tayeb **Mohamed Bouazizi** (29 March 1984 - 4 January 2011) was a Tunisian street vendor who set himself on fire on 17 December 2010, which became a catalyst for the Tunisian Revolution and the wider Arab Spring against autocratic regimes*
[2] ***Mark Duggan**, a 29-year-old British man, was shot and killed by police in Tottenham, North London on 4 August 2011. The Metropolitan Police stated that officers were attempting to arrest Duggan on suspicion of planning an attack and that he was in possession of a handgun. Duggan died from a gunshot wound to the chest*

express their grievance for change or better future. (***Brown et. al., 2013***) *that uprising had given birth to tremendous hopes that the region might see the forging of a new politics—a politics in which those wielding power would find themselves held accountable by the people acting through regular free elections*. The social movement' emotion have been studied by researchers. (***Van Troost et. al., 2013***) *Politics - and especially politics of protest—are full of emotions. People are angry about austerity measures, thrilled or fearful about the Arab Spring and indignant because they want real democracy now!* She suggested that most of the social movement will have some sort of emotions involved either how the protestors own feeling, or the organizer manipulate the emotion of the public. Van Troost et. al. also mentioned in her paper, "*Organizers of protest do not feel constrained by whatever paradigmatic shift. 'Emotion work' has always been a key to the organization of protest. Take the following quote from Malcolm X: "Usually when people are sad, they don't do anything. They just cry over their condition. But when they get angry, they bring about social change." Malcolm X points to an important characteristic of emotions; that is, that emotions propel behavior, but perhaps even more important, different emotions propel different behavior*" (***Van Troost et. al., 2013***).

In this proposal, Social Movement – **Black lives Matter** in The United Stated will be studied. Black Lives Matter is the social movements that has been awhile in America. It is the social movement about the grievance of the African American for equal treatment against racism. The recent incident was trigger by the Death of George Floyd [3](African American) who was killed by the police in Minneapolis, Minnesota on 25 May 2020. Social media platform – twitter data will be used and analyst the sentiment of the public and map it to the social movement cycle by stages, to have better understanding of the mood or emotion changed of the public toward the incident over time along the social movement.

2. Background of The Study

Social Movement Cycle

Helmes-Hayes, in his book he referred to Dawson[4] and Getty's[5] publication in 1929, the social unrest movement can be divided into 3 primary stages - "Collective Excitement", "Formal Organization" and "Institutionalization" (***Helmes-Hayes et. al., 1994***). A paper in Spring 1987 and Mayor, B et. al. proposed there are 8 stages in the social movement (Figure 2.1), where the stages are "Normal Times", "Prove the Failure of Institutions", "Ripening Condition", "Social Movement Take-Off", "Identify Crisis of Powerlessness",

---

[3] ***George Floyd*** *Jr. was an African American man killed during an arrest after a store clerk alleged, he had passed a counterfeit $20 bill in Minneapolis. A white police officer named Derek Chauvin knelt on Floyd's neck for a period initially reported to be 8 minutes and 46 seconds*

[4] ***Carl Dawson****'s development as a sociologist reflected a general trend in sociology's evolution out of theology and social work. Trained as a minister, Dawson rejected the religious vocation at some point after World War I to become a social scientist. Appointed to McGill in 1922, he strove to establish research as the foundation for understanding society, questioning the efficacy of social reform.*

[5] ***Warner Gettys****, American sociologist. Past president Texas Society Welfare Association with Medical Corps, Army of the United States, 1917-1919.*

"Majority Public Support", "Success", and last but not least "Continuing the Struggle" (***Moyer, B et. al. 1987***). The 8 stages from Moyer break the social movement into smaller stages from to Dowson.

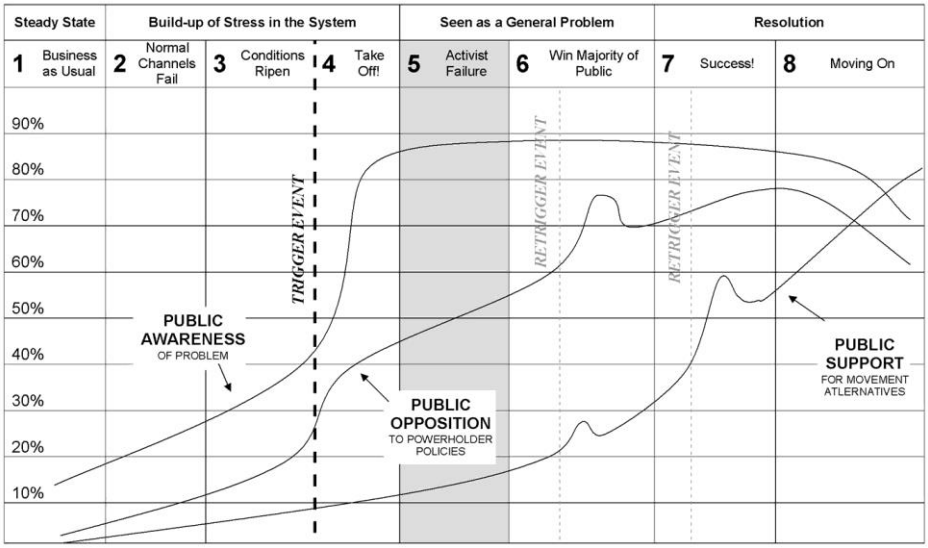

Figure 2.1: Mayor's 8 Stages of the Social Movement

1994, ***Tarrow, S[6] et. al. (1993)*** suggested the theory of Cycle of Protest where it shows the end to end Cycle of the beginning of the protest via each different stage until it ends. However, in 2011 he released another paper said that the individual cycle in the campaign is hard to predict but it is much easier to identify for the beginning and the end. In 2018, a paper published by ***Shultziner, D[7] et. al. (2018)*** further the studied toward the stages and proposed it as "Origin", "Protest", and "Outcome". In his publication, he studied few recently social unrests and mainly focus on the different phenomenon and social behavior for each stage.

Diagram 1: Stage Progression and Causal Mechanisms in Mass Mobilization

Origins		Protest				Outcome	
Grievances	Motivation	Forms of pressures	Degree of organization	Psychological Factors	External forces	Key Players	Strategy
Short-term	Moral shock & anger	Public opinion	No organization	Self-identity	State actions	Political targets	Movement Strategy
Escalating	Inspiration & hope	Disruptive	Organized campaigns	Self-efficacy & empowerment	Counter movements	Independent institutional players	Adversary strategy
Long-term	Organization-generated			Emotions	Media coverage		
Time axis							

Figure 2.2: Shultziner's Stages of the Mass Mobilization

[6] ***Sidney George Tarrow*** *(born 1938) is an emeritus professor of political science, known for his research in the areas of comparative politics, social movements, political parties, collective action and political sociology.*

[7] ***Doron Shultziner*** *is a Visiting Lecturer in the Political Science Department, Gilo Center for Citizenship, Democracy, and Civic Education at Hebrew University of Jerusalem*

Different researchers are having own theory to interrupt the stages of the social movement, however, all theory above are based on their observation, literature reviews, hypothesis and assumptions. As mentioned above, identify the mood or social movement cycle is not easy because for different period of time people behave different toward some topic, not only the individual feeling, the involvement of the government force to suppress the social movement might change the social movement cycle too.

Click here to view the complete literature review for the proposal.

Word Level Lexicon

F. S. Tabak et al (2016) published a paper to compare EmoSenticNet (**ESN**), National Research Council Canada (**NRC**), DepecheMood (**DPM**), and Topic based DepecheMood (**TDPM**). According to his conclusion, the NRC and DPM have better classification results.

Emotion / Sentiment Analysis

In this project, the proposal is not to do the polarity of the sentiment (Positive, Negative, and Neutral), but will study the fined-grain emotion by using ***Plutchik, R's*** method. By referring to the Figure 2.4 below, there are 8 major moods which are "joy", "trust", "fear", "surprise", "sadness", "disgust", "anger", and "anticipation".

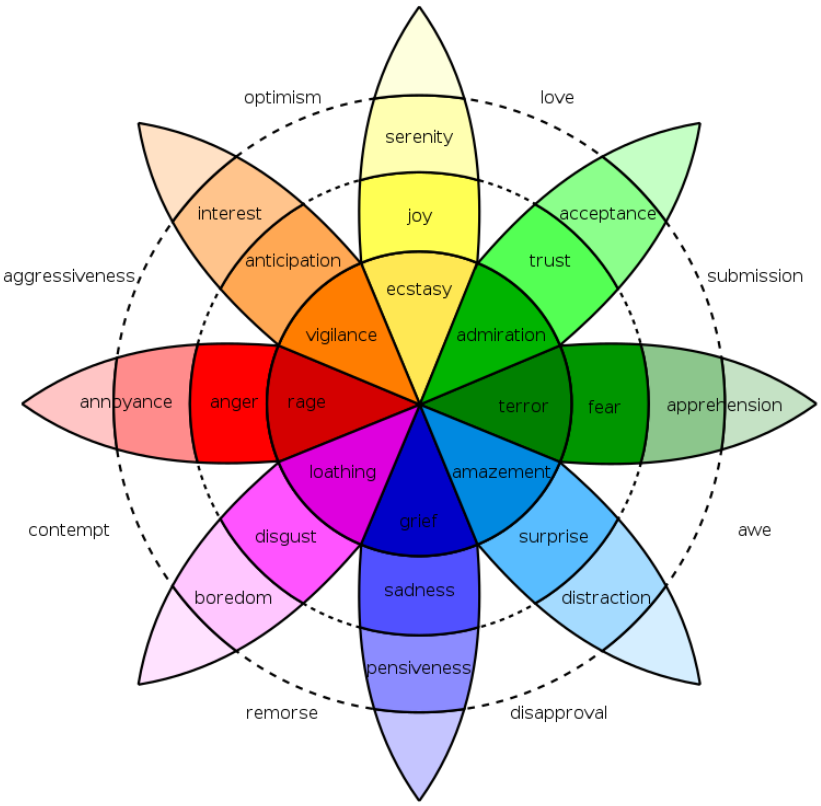

Figure 2.4: Plutchik Wheel of Moods

For this project, Deep learning technique will be used to do the classification on the tweets data to label it with one of the 8 moods according to Plutchik's. Text data (for example tweets) will be collected, and data cleaning will be done (Refer to Methodology below) and Machine Learning Approaches will be used to classify the mood. In ***Sundari, A et al. (2020)*** paper, he performed study on few different deep learning methods like CNN, LSTM and BI-LSTM on text. He claimed that the Bi-LSTM has much better results (0.685 Accuracy) for each emotion compared to other. Another similar paper from ***Chatterjee, A et. al.*** published in 2019 for SemEval 2019 conference and LSTM and Glove was being used, they able to get 0.7959 F1 score. ***Agrawal et. al. (2019)*** also published a paper from the same SemEval 2019 conference. In this paper, they used a method name NELEC which is the combination of Lexical and Neural Network method, and **LSTM** is the Neural Network method in their project. 0.7765 is the F1 score.

Time Series

After the time series data by day generated from the sentiment analysis. Statistic methods needed to measure the time series similarity. There are few popular and widely use methods for example Dynamic Time Warping (**DTW**), Auto Regressive Integrated Moving Average (**ARIMA**), and Pearson Correlation methods. For example, ***A.C. Linke et al. (2020)*** compared DTW and Pearson Correlation to compare the frequency from Resting State fMRI. From 5 different analyses perform by the researcher, they concluded that the DTW performs at superior level compared to Parson Correlation.

3. Research Issues / Problem Statement

Understanding Social Movement is not an easy task, where it consists of many possible factors. For a decade, researchers have been studying it from various point of view, like the social science, psychology, as well as data - to predict the start of the social movement. The topic remains challenging, due to the changes from the environment for example, Social Media that spread the information faster and specific topic being objectify and manipulate by some parties to influence the public opinion. Especially recently year more and more social movements occurring all over the world for various reason. Understanding the social movement emotion change motif is some interesting topic to be studied. By understanding the emotions along the social movement cycle, relevant parties can plan for the actions for negotiation and avoid the social movement escalated to another level. By making sense of the cycle it can be the baseline for other researcher to further the study on social movements, like predict the start, end, or even the change of the stage of the social movement cycle.

4. Research Questions

To have better understand how the public sentiment toward Social Movement. e.g. Black Lives Metter, London Riots, Yellow Vest Riot, and Hong Kong Umbrella. How and when the mood change and the pattern of it.

5. Research Objectives

Hypothesis: Regardless of time and location, mood change motif during the social movement cycle are the same.

6. Methodology

Following methods are needed to accomplish the hypotheses above. The methods are divided into 3 parts.

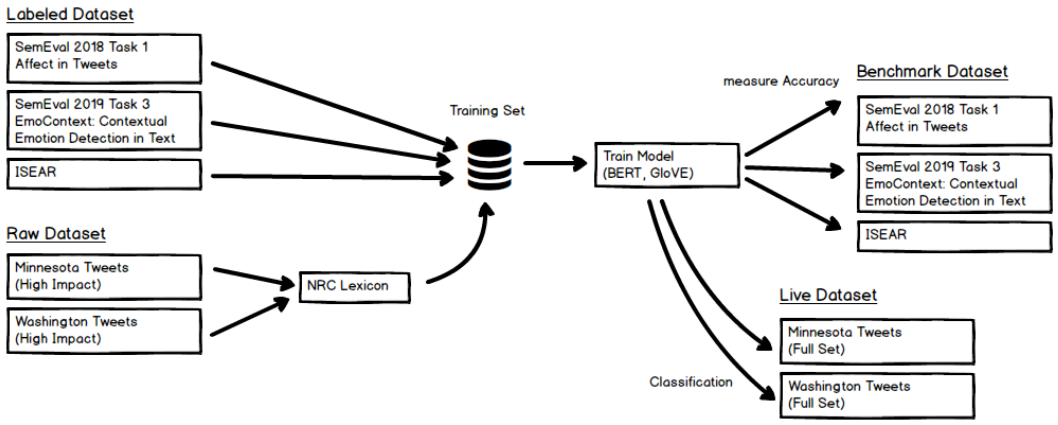

Figure 6.1: Part A Methods

Part A: Data

Two set of datasets will be used in this project, 3 set of labelled datasets and 2 set of unlabeled datasets. Both 2 set of datasets will then become the training dataset to train the model for mood classicization. Please refer to the Figure 6.1 above.

Labeled datasets

3 set of labelled datasets will be used in this project to train the classification model. Those datasets have been used by other researchers and it has the mood labeled needed by this project. However, the mood set for each dataset are different. Hence in this project, we need to adjust and combine the mood into the standard proposed by Robert Plutchik (2001) [13]. Table 6.1 below shows the number or records for each mood and map it to Plutchik'

standard. ***Preliminary Test*** executed and result is available in these links, SemEval-2018, SemEval-2019, ISEAR.

Plutchik's	SemEval 2018	SemEval 2019	ISEAR	Total	Remark
Fear	1,242		1,074	2,316	All dataset in the 8 Plutchik's mood will be accepted as training data.
Anger	2,544	5,506	1,079	9,129	
Sadness	2,008	5,463	1,082	8,553	
Trust	357			357	
Joy	2,477	(Happy) 4,243	1,094	7,814	
Surprise	361			361	
Anticipation	978			978	
Disgust	2,602		1,066	3,668	
	(Love) 700				Discard
	(Optimism) 1,984				
	(Pessimism) 795				
		14,984 (Others)			
			1,071 (Shame)		
			1,049 (Guilt)		

Table 6.1: Part A Method

Unlabeled datasets

Unlabeled datasets are the raw data downloaded from twitter by using GetOldTweets API by Keyword #BlackLivesMatter and date from May 2020 to July 2020. In order to build the training data from this raw data, only the tweets with high rewet and favorite count will be used. The raw data then will be label with mood by using NRC Lexicon and the intensity will be count.

```
final_tweets = []
for tweet in tweets:
   if tweet.retweet > X AND tweet.favorite > X:
      if tweet.intensity_score > Z:
            final_tweets.append(tweet)
      end if
   end if
end for
```

Pseudo Code 1: Filter the Unlabeled Data

The X and Y value in the Pseudo Code 1 above are the number of minimum retweet count and favorite count. For the Z value the minimum intensity score. In order to define a food minimum Z value, the bell curve graph for each mood intensity will be populated by looking into the standard deviation, medium and mean value to define the Z value. Only the mood instance the above Z (e.g. 75%) above will be used.

Figure 6.2: Bell Curve to determine the Z value

Combining the labeled and unlabeled dataset above will have high chance to get imbalance class. Hence the Cost Function Method need to be considered to handle the imbalance class. ***Preliminary Test*** executed and result is available here.

Word-Level Sentiment Classification

Due to some datasets are unlabeled, the Word-Level Lexicon needed to label the mood and the intensity. According to the previous related work by ***F. S. Tabak et. al.*** (2016). He concluded the National Research Council Canada (**NRC**) and DepecheMood (**DPM**) have better result in term of classification. Simple preliminary tests have been executed on the 2 set of Labelled datasets to measure the accuracy. Click these link to see the complete results (1, 2, and 3) . Below are the results on one sample dataset (ISEAR) by using NRC, DepecheMood++ and EmoSenticnet. Please refer to the 3 figures below (6.3, 6.4, and 6.5), the yellow color represents the number of the instances in ISEAR dataset. And the color bars represent the number of instances classified with the mood. For example, the Figure 6.3 below, the total instances per mood class (Yellow Box) in ISEAR are quite balance, which is approximately 1000 instance per class. However, the EmoSenticNet labelled majority of the instance as "Joy" which is incorrect. Only the "Sadness" mood appear to be slightly better than other.

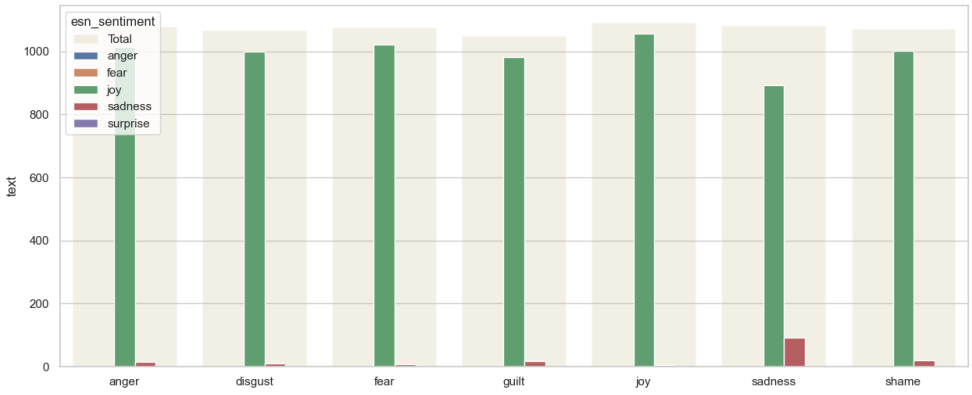

Figure 6.3: EmoSenticNet Lexicon on ISEAR Dataset

Figure 6.4 below is the result labeled by DepecheMood++ Lexicon, as you can see the Pink bar are the majority for most of the mood class. Only the "Angry" and Sadness" have better accuracy but anyhow it is still performed poorly.

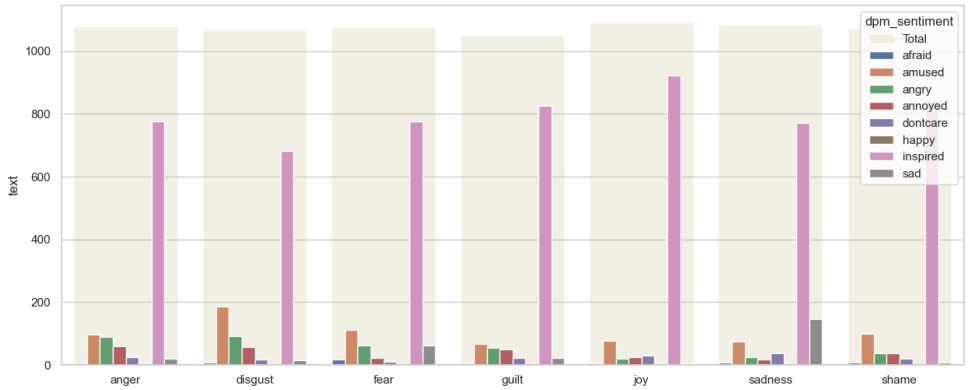

Figure 6.4: DepecheMood++ Lexicon on ISEAR Dataset

Figure 6.5 (by NRC) are not great but better among all 3. As you can see from the figure below it has approximately 20-30% accuracy.

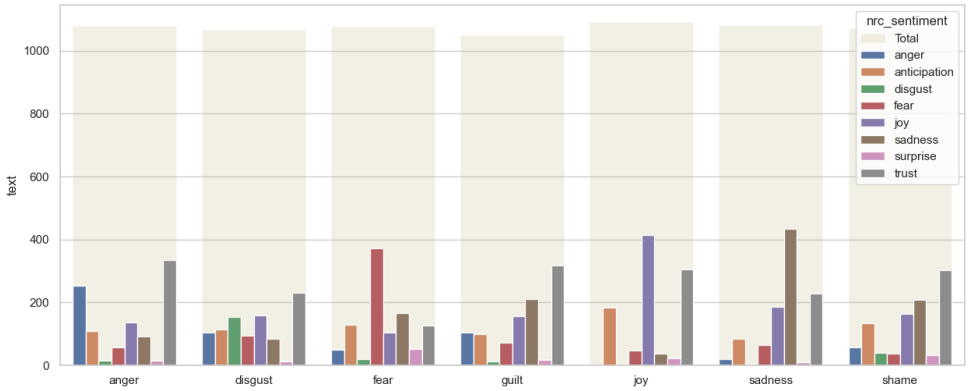

Figure 6.5: NRC Lexicon on ISEAR Dataset

In conclusion, NRC can perform better than EmoSenticNet and DepecheMood++, only need to fine tune the word-level classification.

Train Model, Accuracy, and Classification

In order to build the Model, several tasks need to be done. First the data will need to be cleaned and following are the tasks

- Detect the Language of the tweets, only keep the English language.
- Spelling correction and replace missing word.
- Remove @USERNAME, URL
- Sprit the hashtag to word. E.g. "#IHateYou" to "I Hate You"
- Remove Punctuation and Symbol

- Lemmatize and Tokenize the tweet

Glove / BERT will be used as word embedded, use Deep Learning technique (LSTM) to train the model for further mood classification, Cost Function Method need to consider if class imbalance issue. 3 labeled datasets will be used to measure the accuracy of the model and final model will be execute on the full set of unlabeled / raw tweets data to classify the mood for each individual tweet.

Part B: Descriptive Analysis for the Social Movement Cycle

The classified data from Part A above will be used to extract the number of tweets count and the mood intensity by day. This part it to assess the ***hypothesis I*** where the mood change motif for 2 locations should be identical. Following is the Illustration of the line chart for 2 locations.

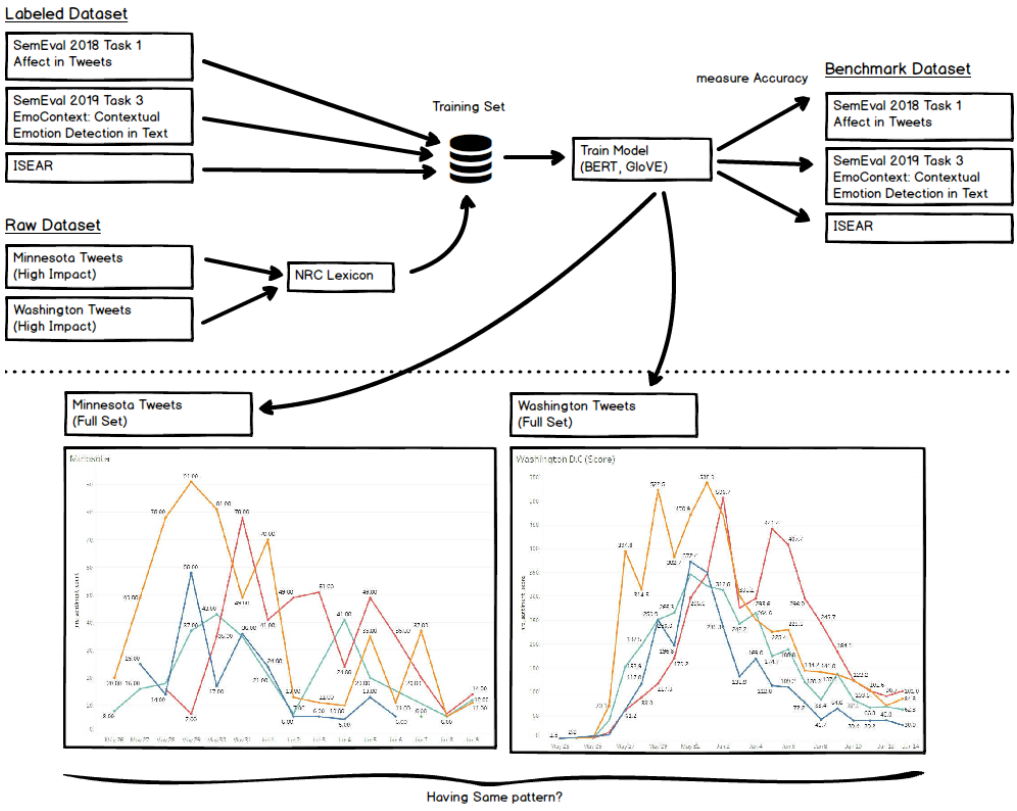

Figure 6.3: Illustrated Line Chart (Number of Tweet per Day)

Preliminary was done on the dataset by using Lexicon Approaches (NRC) to have brief understanding on the data. Following Figure 6.4 shows the sample data visualized by Tableau, it shows the line chart of number of tweets per day by moods. Although in general the shape of the curves are not 100% same, but it has the similarity there.

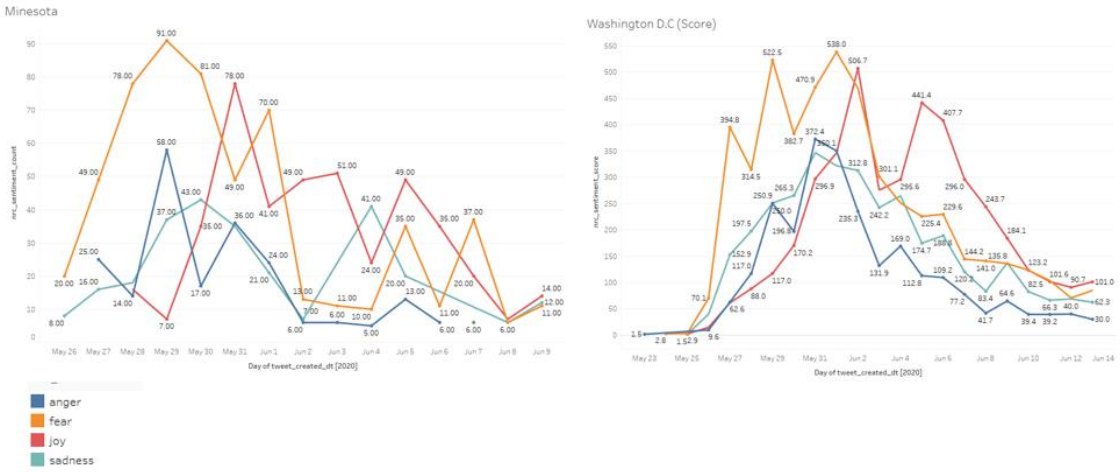

Figure 6.4: Trend line comparison (Removed Trust)

For example
- *Orange* ***(Fear)*** comes into picture first, then blue ***(Anger)*** and Green ***(Sadness)***, and Last Red ***(Joy)***
- Red ***(Joy)*** line becomes higher than Green and Blue at some point (20% of the cycle).
- Blue ***(Anger)*** line lowest at some point (40% of the cycle)

Each of the mood will be extract from each location, for example the fear from Location A and location B. **DTW** (Dynamic Time Warping) method will be used to compare both time series (in day). Figure 6.4 shows that the distance matrix will be computed from two time series in order to compare the similarity.

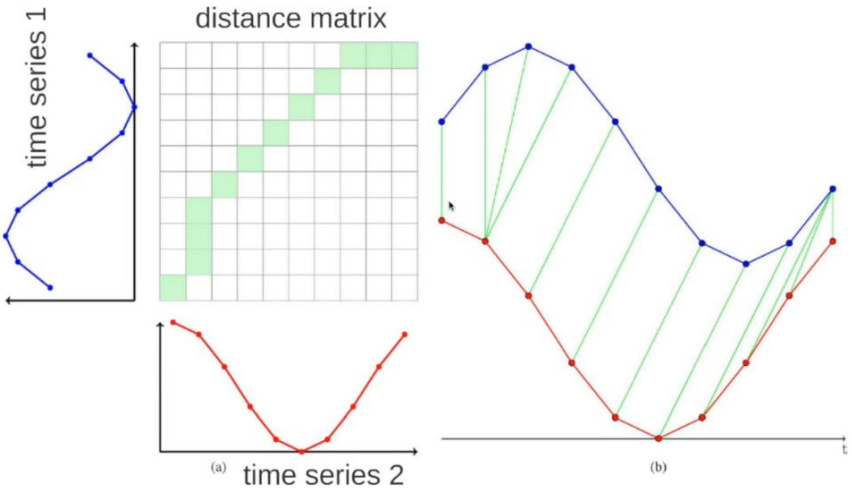

Figure 6.5: Distance Matrix Concept in Dynamic Time Warping

7. Research Scopes

This study aims to understand more about the emotions of the public during social movement. Following techniques are the major focus in this research, which are the **Hybrid** Approaches for **Sentiment** Analysis, **Time Series**, and **Statistic Factor Analysis**. Latest **Word Embedded** – BERT will be explored.

The limitation of this research is (1) not able to study all locations with Black Lives Matter social movement. It is because the Twitter API changed and the GetOldTweets API not able to pull more data. 2nd limitation (2) is only focus on English tweets, to avoid the needs of translation that might potentially reduce the accuracy of the study.

8. Significance of Study

The finding of this study will able to prove the theory from the social science (Collective Behavior and Social Movement Stages) by the empirical study on the tweet data. Besides, by understanding the mood change motif, can help the further study to understand the social movement stages and impose proper action to avoid the escalation of the incident. Identify the impact factor of the mood toward the police cases can also help to find out the type of mood drive the public to involve in the crimes. Based on the finding, prevention action can be introduced.

9. Literature Review

A.C. Linke, L.E. Mash, C.H. Fong, M.K. Kinnear, J.S. Kohli, M. Wilkinson, R. Tung, R.J. Jao Keehn, R.A. Carper, I. Fishman, R.-.A. Müller, Dynamic time warping outperforms

Pearson correlation in detecting atypical functional connectivity in autism spectrum disorders, NeuroImage, Volume 223, 2020, 117383, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2020.117383.

Agrawal, Parag, and Anshuman Suri. “NELEC at SemEval-2019 Task 3: Think Twice Before Going Deep.” Association for Computational Linguistics (ACL), 2019. doi:10.18653/v1/s19-2045.

Albalawi, Rania, Tet Hin Yeap and Morad Benyoucef. “Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis.” Frontiers in Artificial Intelligence (2020).

Brown, Nathan. (2013). Egypt's Failed Transition. Journal of Democracy. 24. 45-58. 10.1353/jod.2013.0064.

Chatterjee, A., Kedhar Nath Narahari, Meghana Joshi and P. Agrawal. “SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text.” SemEval@NAACL-HLT (2019).

F. S. Tabak and V. Evrim, "Comparison of emotion lexicons," 2016 HONET-ICT, Nicosia, 2016, pp. 154-158, doi: 10.1109/HONET.2016.7753440.

Helmes-Hayes, Richard C. “Canadian Sociology's First Textbook: C. A. Dawson and W. E. Gettys's ‘An Introduction to Sociology (1929)".” The Canadian Journal of Sociology / Cahiers Canadiens De Sociologie, vol. 19, no. 4, 1994, pp. 461–497. JSTOR, www.jstor.org/stable/3341148. Accessed 17 Oct. 2020.

I. Vayansky and S. A. P. Kumar, "A review of topic modeling methods", Inf. Syst., vol. 94, Dec. 2020

J. Qiang, Z. Qian, Y. Li, Y. Yuan and X. Wu, "Short Text Topic Modeling Techniques, Applications, and Performance: A Survey" in IEEE Transactions on Knowledge & Data Engineering, vol. , no. 01, pp. 1-1, 5555

Lewis, Paul, Newburn, Tim, Taylor, Matthew, Mcgillivray, Catriona, Greenhill, Aster, Frayman, Harold and Proctor, Rob (2011) Reading the riots: investigating England's summer of disorder. Reading the riots. The London School of Economics and Political Science and The Guardian, London, UK.

Moyer, B. (1987). The movement action plan: A strategic framework describing the eight stages of successful social movements. Retrieved from http://www.historyisaweapon.com/defcon1/moyermap.html

Plutchik, R. (2001). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist, 89(4), 344-350. Retrieved October 18, 2020, from http://www.jstor.org/stable/27857503

Tarrow, S. (1993). Cycles of Collective Action: Between Moments of Madness and the Repertoire of Contention. Social Science History, 17(2), 281-307. doi:10.1017/S0145553200016850

Shultziner, D, Goldberg, S. The stages of mass mobilization: separate phenomena and distinct causal mechanisms. J Theory Soc Behav. 2019; 49: 2–23. https://doi.org/10.1111/jtsb.12187

Sundari, A. and R. Shenbagavalli. "A COMPARATIVE ANALYSIS OF EMOTION PREDICTION ON A TEXT BY CNN, LSTM AND BI-LSTM." Journal of emerging technologies and innovative research (2020)

Van Troost, Dunya & Stekelenburg, Jacquelien & Klandermans, Bert. (2013). Emotions of Protest. 10.1057/9781137025661_10.

Ye Yiran and Sangeet Srivastava. 2019. Aspect-based Sentiment Analysis on mobile phone reviews with LDA. In Proceedings of the 2019 4th International Conference on Machine Learning Technologies (ICMLT 2019). Association for Computing Machinery, New York, NY, USA, 101–105. DOI:https://doi.org/10.1145/3340997.3341012