

Measuring Online Emotional Reactions to Events

Siya Guo¹, Zihao He¹, Ashwin Rao¹, Eugene Jang², Yuanfeixue Nan²,
Fred Morstatter¹, Jeffrey Brantingham³, and Kristina Lerman¹

¹Information Sciences Institute, ²University of Southern California, ³University of California, Los Angeles
{siyiguo, zihaohe, mohanrao, eugeneja, ynan}@usc.edu, fredmors@isi.edu, branting@ucla.edu, lerman@isi.edu

Abstract—The rich and dynamic information environment of social media provides researchers, policy makers, and entrepreneurs with opportunities to learn about social phenomena in a timely manner. However, using this data to understand social behavior is difficult due to heterogeneity of topics and events discussed in the highly dynamic online information environment. To address these challenges, we present a method for systematically detecting and measuring emotional reactions to offline events using change point detection on the time series of collective affect, and further explaining these reactions using a transformer-based topic model. We demonstrate the utility of the method on a corpus of tweets from a large US metropolitan area between January and August, 2020, covering a period of great social change. We demonstrate that our method is able to disaggregate topics to measure population’s emotional and moral reactions. This capability allows for better monitoring of population’s reactions during crises using online data.

Index Terms—emotional reaction, social media data, change point detection, topic modeling

I. INTRODUCTION

Social media platforms connect billions of people worldwide, enabling them to exchange information and opinions, express emotions, and to respond to others. Researchers, policy makers, and entrepreneurs have grown interested in learning what the unfettered exchange of information reveals about current social conditions, including using social media data to track public opinion on important issues [1] and monitor the well-being of populations at an unprecedented spatial scale and temporal resolution [2].

Using social media data to learn about human behavior, however, poses significant challenges. Social media represents a heterogeneous, highly dynamic information environment where some topics are widely discussed while others are barely mentioned [3]. It includes people’s self-reports of their own lives, as well as reactions to external events. Researchers have developed methods to detect events from online discussions, including clustering text based on similarity, analyzing term co-occurrence, identifying bursty terms and deep learning techniques [4]–[7]. However, social media data provides evidence for learning about human behavior beyond shifts in

topics. For example, it can also shed light on emotions and morality, which are important drivers of individual attitudes, beliefs, and psychological and social well-being [8], [9].

To study the collective affect, researchers investigated how social media content influences emotional user engagement [10], [11]. These works, however, leave a gap in our understanding of collective emotional and moral reactions to socio-political events, which could shed light on opinion dynamics, emergence of polarization, and help identify online influence campaigns.

To bridge these gaps, we present a methodology for detecting, measuring and explaining the collective emotional reactions to offline events. Using state-of-the-art transformer-based models, we construct the time series of aggregate affect from social media posts. We detect emotional reactions as discontinuities in these time series, and then explain the reactions using topic modeling. We demonstrate the utility of the methodology on a corpus of tweets collected from a large US metropolitan area between January and August, 2020. This time span represents a complex period in American history with important social, political and cultural changes. We successfully detect the simultaneous crises of the COVID-19 pandemic and racial justice reckoning, and other important events like political primaries. We show how these developments had profound impact on the psychological state of the population. For example, as the COVID-19 pandemic began to unfold, people expressed more anger and fear, and more moral sentiments like care and authority. Furthermore, we disaggregate COVID-related tweets by topic to more accurately measure the emotional reactions to different subtopics. We identify stronger reactions to daily-life issues (e.g. grocery panics) than topics directly mentioning COVID-19.

While we perform analyses on Twitter data, our pipeline is generalizable to other social media platforms and news. Our results suggest that studying the collective emotional reactions on social media can provide valuable insights into understanding people’s opinions and responses to timely socio-political events, and aid policy makers in crafting messages that align with the values and concerns of the population¹.

II. METHODS AND MATERIALS

To understand the dynamics of affect, we propose a pipeline (Fig. 1) that detects, measures and explains online emotional

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ASONAM '23, November 6-9, 2023, Kusadasi, Turkey

© 2023 Copyright is held by the owner/author(s).

ACM ISBN 979-8-4007-0409-3/23/11

<http://dx.doi.org/10.1145/3625007.3627477>

¹See the full version of this work [here](#).

and moral reactions to offline events. With a set of time-stamped texts, e.g. tweets, we first perform emotion and morality detection from text. We then construct the time series of the aggregate affect on a daily basis. Next, to detect reactions, we perform change point detection on each emotion and morality time series. We measure the magnitude of the change at each detected change point and perform topic modeling to explain the offline event that triggered the specific online reaction.

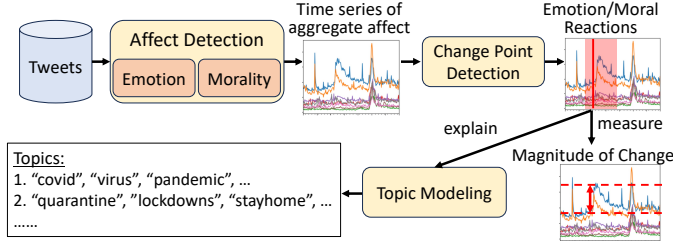


Fig. 1. Pipeline to detect and measure online emotional reactions.

A. Data

The data used in this study was collected using Twitter’s Filter API by specifying a geographic bounding box over a large metropolitan area. This method collects every tweet that is either geotagged within the bounding box (using the device’s coordinates with the user’s permission), or by using the Twitter “place” feature, where the user tags their location. We collected 17M tweets from 350K unique users.

B. Emotion and Morality Detection

We first measure emotions and moral sentiments expressed in an individual tweet. For emotion detection, we use a state-of-the-art language model SpanEmo [12], fine-tuned on the SemEval 2018 1e-c data [13]. This transformer-based model outperforms prior methods by learning the correlations among the emotions. It measures *anticipation*, *joy*, *love*, *trust*, *optimism*, *anger*, *disgust*, *fear*, *sadness*, *pessimism* and *surprise*.

We quantify the moral sentiments of tweets along five dimensions [9]: dislike of suffering (*care/harm*), dislike of cheating (*fairness/cheating*), group loyalty (*loyalty/betrayal*), respect of authority and tradition (*authority/subversion*), and concerns with purity and contamination (*purity/degradation*). We fine-tune a transformer-based model on diverse training data (see [14] for details). The large amount and the variety of topics in our training data helps mitigate the data distribution shift during inference. After labeling tweets, we calculate the daily fractions of tweets with different emotion and moral categories to construct the time series.

We evaluate the effectiveness of emotion and morality detection on a random subset of 850 tweets, considering high difficulty and time cost of the annotation task. We asked five educated annotators to go through two training sessions, in each they annotated 50 random tweets and discussed to improve the agreement on the definitions of emotions and morality. Then each annotator individually annotated all 850 tweets. The Fleiss’s κ for emotion categories ranges in

0.42 ± 0.02 . For morality categories, it ranges in 0.30 ± 0.03 . Emotion and morality are very subjective concepts. Similar to prior works [13], [15], we have found the κ scores of some categories low. However, our agreement is still comparable and even better than these prior works.

On this annotated dataset, we compare our emotion and morality detection methods with widely used dictionary-based methods, namely keyword matching using Emolex for emotions, and Distributed Dictionary Representations (DDR) [16] for morality. Our methods outperforms baselines on ten out of 11 emotion categories (the F1-scores of our method are in 0.42 ± 0.14 , and those of Emolex are in 0.15 ± 0.11) and we outperform on nine out of ten moral categories (the F1-scores of our method are in 0.31 ± 0.17 , and those of DDR are in 0.17 ± 0.15). The performance inevitably varies with support for different categories, as also observed in previous studies [15]. Despite the variation in model performance, prior research [2] has validated that when aggregating on the collective level, the time series of sentiments constructed with supervised deep learning detection and dictionary-based methods have strong correlations with those from self-reports.

C. Detecting and Measuring Change Points

The time series of emotions and morality reveal the complex dynamics of aggregate affect on social media. We define an emotional reaction as a change in the corresponding time series. To detect such change points, we combine two popular methods. The first, cumulative sum (CUSUM) method [17], detects a shift of means, and is good at detecting changes like the COVID-19 outbreak, which shifted the baseline emotion and moral sentiment. To detect multiple change points, we use a sliding window to scan the whole time series. We set the window size to be four weeks and slide it every five days for the best precision. Another type of event, such as Valentine’s Day, creates a short surge of emotions, can be better detected with Bayesian Online Change Point Detection (BOCPD) [18]. It uses Bayesian inference to determine if the next data point is improbable, which is good at detecting sudden changes. We identify a change point to be significant when either CUSUM or BOCPD gives a significant confidence score, using 0.5 as threshold. We perform change point detection separately for each time series of emotion and morality, because different types of events may elicit different reactions.

For each detected change point, we quantify the magnitude of the collective reaction as percent change before and after it. We compute the baseline level before the change point as the mean of the time series over the two week period before. Then, we measure the short-term and long-term changes. To calculate the short-term change, we compare the baseline to the peak or dip value in the two weeks after the change point and compute percent change. To calculate the long-term change, we compare the baseline to the time series value two weeks after the event (we take a five-day average around the two-week mark). The size of the window is empirically chosen to be two weeks so that enough observations are made, but it would not be affected by another event earlier or later.

D. Explaining Changes with Topic Modeling

We try to explain changes in emotions detected by our method using topic modeling. We choose BERTopic [19], a transformer-based language model that extracts highly coherent topics compared to traditional LDA. We evaluate both methods on a set of 10% randomly selected tweets from our data, using a different numbers of topics ranging from 10 to 50 in steps of 10. Over different runs, BERTopic gives higher NPMI coherence scores (0.14 ± 0.01) compared to LDA (0.03 ± 0.01), and similar diversity [20] scores (0.75 ± 0.04) compared to LDA (0.76 ± 0.04).

For each emotional reaction, we extract the topics of tweets that are tagged with that emotion or morality category. We apply BERTopic to tweets within the three-day time window before and after the change, as discussions quickly die on social media [21]. For example, for the Black Lives Matter (BLM) protests starting on 2020-05-26, we extract the topics from tweets posted between 05-23 to 05-25 to develop a baseline and then separately extract the topics between 05-26 to 05-28. By comparing the top 10 baseline topics before the change point with those after the change point, we determine the new topics that emerged after the change points that are possibly relevant to the event.

For preprocessing, we remove URLs and name mentions, transform emojis to their textual descriptions, and split hashtags into individual words. We use the Sentence-BERT “all-MiniLM-L6-v2” model [22] to directly embed the processed tweets. After topic modeling, we remove English stopwords in the learned topic keywords. With each emerging topic, we manually verify if there is an associated offline event by examining the tweets belonging to this topic and by searching related news articles. Such manual verification is a necessary and common practice event detection literature [6].

III. RESULTS

A. Online Reactions to Offline Events

Time series of the aggregate affect from January to August 2020 (Fig. 2) shows complex dynamics with seasonal variation (weekly cycles in joy), short-term bursts (spike in love on Valentine’s Day), and long-term changes in emotions and moral sentiments. This time span represents a difficult period in the life of the city. In addition to the world-wide pandemic, which led to a national lockdown mid-March, political primaries were also taking place during this time period, which also saw one of the largest social justice protests triggered by the murder of George Floyd in police custody, as well as the death of a beloved sports icon. These developments had a profound impact on the city’s population, as demonstrated by the many rises and dips in emotions and moral sentiments.

We ran the proposed pipeline to detect and explain the online emotional reactions to events. Table I shows that our method is able to identify larger and impactful events such as the COVID-19 pandemic and the BLM protests. We see the complex reactions to the pandemic in multiple dimensions of emotions and morality. The unsupervised method also enables

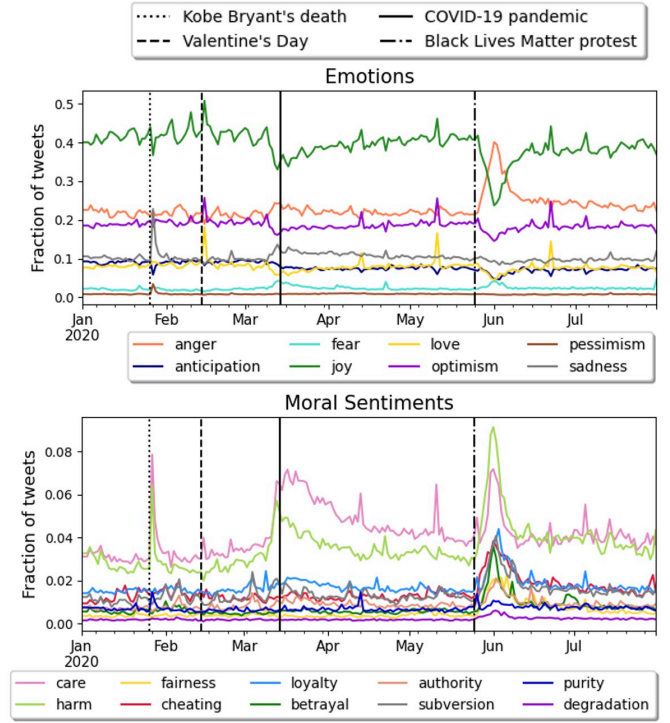


Fig. 2. Time series of emotions and moral sentiments from January 1 to August 1, 2020. We show the daily fraction of tweets with different affect labels. The notable peaks and dips in the time series can be associated with the external events marked as vertical lines.

us to discover reactions to smaller events that might be easily missed, such as earthquakes and baseball playoffs. We also show that running BERTopic on tweets posted near the event reveals the relevant topics. Further, because we detect changes separately in each emotion, we can disentangle events based on different emotional reactions, even when they take place on the same day: e.g., Trump’s impeachment trial was associated with an increase in betrayal and subversion, MLK Day with joy, loyalty and fairness, and earthquake with fear.

B. Evaluation of the Proposed Pipeline

Similar to prior work [6], [23], we use precision and duplicated event rate (DERate) to evaluate our method. Recall is not used because we cannot annotate every tweet to obtain an exhaustive list of events. Precision is the fraction of detected events that are related to realistic events [24]. We manually verify each event by searching news with topic keywords associated with each change point, which is common practice in event detection research. We detected 54 change points in total, with confidence ranging from 0.65 to 1.00, and 85% with confidence score above 0.9. We found 10 false positive, which cannot be explained by any topic and/or be related to any real event, giving a precision of 0.84.

Duplicate Event Rate (DERate) is the percentage of duplicate detected events among all realistic events detected [23]. We define it as the fraction of emotion and morality categories (out of 21) that detected the same event. The higher DERate

TABLE I
TOP EVENTS AND THE EMOTIONAL AND MORAL REACTIONS DETECTED BY OUR METHOD, IN DESCENDING ORDER OF IMPACT.

	Event	Date	Time Window	Peaking Emotion/MF	Declining Emotion/MF	Relevant Topics
1	BLM protests	05-26	05-25 to 05-30	anger, disgust, fear, care, harm, fairness, cheating, loyalty, betrayal, authority, subversion, degradation	anticipation, joy, love, optimism	"racist, racism, white", "murder, george, floyd", "black, injustices, oppresion"
2	COVID-19 pandemic	03-10	03-09 to 03-11	anger, disgust, fear, sadness, care, harm, authority	anticipation, joy, love, optimism	"coronavirus, corona, virus", "pandemic, virus, administration", "safe, stay, everyone"
3	Kobe Bryant's death	01-26	01-26	anger, pessimism, sadness, care, harm, purity	joy	"kobe, rest, peace", "prayers, praying, family"
4	Valentine's Day	02-14	02-14	love	anger, disgust	"valentine, valentines, happy"
5	MLK Jr. Day	01-20	01-17 to 01-20	joy, loyalty, fairness		"mlk, king, martin", "rights, king, justice"
6	Trump's State of the Union address	02-03	02-03	anger, disgust, authority		"pelosi, nancy, speech", "pelosi, trump, tear"
7	Trump's impeachment trial	01-20	01-20 to 01-21	betrayal, subversion		"senate, vote, will", "treason, traitors, moscow"
8	CA Pres. primary	03-03	03-02 to 03-03	fairness, subversion		"vote, california", "biden"
9	DOJ Drops Flynn Case	05-07	05-07	cheating		"flynn, cheated, cheating"
10	Dodgers beat Giants	07-23	07-23	joy		"dodgers, mlb, giants"
11	Earthquake	01-20	01-20	fear		"earthquake, usgs, km"

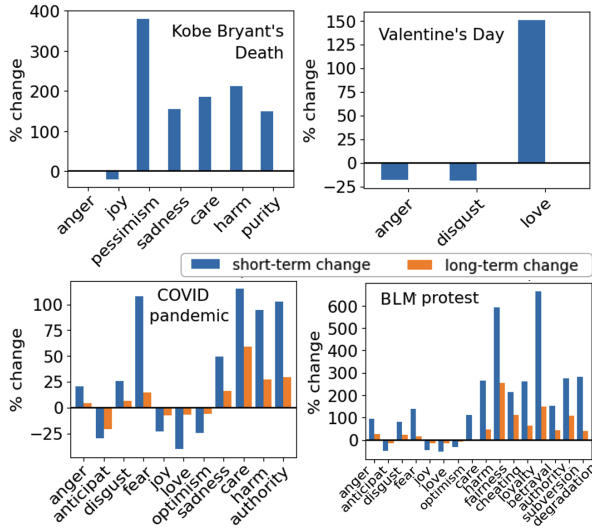


Fig. 3. Short-term and long-term changes of emotions and moral sentiments around four events. The short-term change compares the peak/dip value after an event to the baseline level before the event. The long-term change compares the time series value around two weeks after the event to the baseline level.

shows better confidence. The DERate of our method is 0.14. Our precision and DERate are comparable to prior works [6].

C. Short-term and Long-term Changes in Affect

Our proposed method enables us to study collective reactions to events along multiple dimensions of affect. For example, the BLM protests were associated with 16 different emotional and moral changes. We quantify the percent change in the corresponding collective affect before and after the event for four of the most impactful events (Fig. 3). Consistent with our intuition, Kobe Bryant's Death was associated with a short-term increase in *pessimism* and *sadness* and a decrease in *joy*,

as well as a short-term rise in moral language related to *care* and *harm*. In contrast, Valentine's Day brought a short-term increase in *love* and a decrease in *anger* and *disgust*. No long-term changes were seen with these events.

The COVID-19 outbreak triggered a cascade of events aimed at mitigating the pandemic that were associated with complex short-term and long-term changes in affect. People expressed more *anger*, *disgust*, *sadness*, and more significantly, *fear*, both in the short-term and the long-term. Positive emotions like *joy* and *love* simultaneously decreased. People also expressed more moral sentiments like *care* such as in "Stay safe. We thank you", as well as more *harm* blaming the virus. Interestingly, the moral language around *authority* also increased, possibly due to new policies such as lockdowns to mitigate the pandemic (e.g. "I think governor Newsom is doing a great job..."), and some were critical of government's response, e.g., "we need leadership not a politician".

The BLM protests was also associated with complex short- and long-term changes in affect. We observe increases in negative emotions and decreases in positive emotions. In addition, compared to other three events, we see greater increases in moral sentiments. The moral concerns about *fairness* and *betrayal* had especially increased, expressing a deep sense of the injustice and betrayal in George Floyd's death.

D. Disentangling COVID-related Emotions

The COVID-19 pandemic was associated with complex and long-term emotional changes. Here we use this example to further show the benefit of disentangling emotional reactions by disaggregating topics. We select four top categories discussed and group related topics into these categories: directly covid-related topics, grocery panics, leisure activities and school and education. We study emotions and moral expressions

aggregated in all the tweets, as well as in these topic categories (Fig. 4). We find that aggregating emotions from all tweets can give misleading impressions. Positive emotions like *joy* were highly expressed in all tweets (aggregated), but in fact they were mostly dominated by people talking about leisure activities. In COVID-related tweets, few positive emotions were expressed. *Anger* and *disgust* were higher in topics about grocery panics than in topics directly related to COVID. Another example is the expression of *care* and *harm* moral sentiments. Their expressions were diluted by other topics in aggregate tweets. By disaggregating, we see that they were highly expressed in directly COVID-related tweets. These results suggest that during times of maximal crisis and uncertainty, people find outlets for positive emotions. They also demonstrate the importance of disaggregating by topics when studying specific issues.

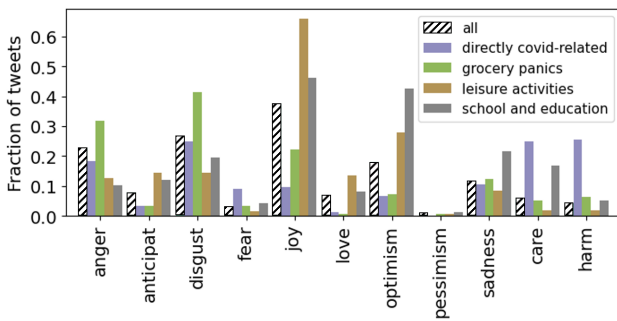


Fig. 4. Emotions and moral sentiments expressed in COVID-related topics during the two weeks after WHO announcement of the pandemic on 2020-03-11. The topics are COVID (“coronavirus, corona, virus”); grocery panics (“grocery, groceries, shelves”, “water, dasani, hydro”, and “toilet, paper, rolls”); leisure activities (“episode, episodes, show”, “cook, cooking, cook-out”, “tickets, ticket, selling”); and education (“teachers, students, learning”, “schools, lausd, classes”, “schools, lausd, closed”).

IV. CONCLUSION

In this work, we have demonstrated the effectiveness of an unsupervised method to detect and measure public reactions to newsworthy events. We applied our method to a large Twitter corpus of tweets drawn from the population of a large metropolitan area, disentangled the dynamics of online emotions during a time period punctuated by complex social, health, and political events. We showed that our method can discover significant and less significant events and measure emotional and moral reactions to these events. To further understand the complex impact of the COVID-19 outbreak, we disaggregated COVID-related tweets and discovered topics directly related to the virus and topics related to changes in life style, including unemployment, grocery panics and education. The emotions expressed on these different topics suggest that people had negative feelings and thoughts during the height of the pandemic, but were also searching for the positive and holding on to optimism. Together, these results suggests the potential of using social media data for tracking of public reactions to events, as well as discovering significant events that may have been missed by traditional news sources.

There are some limitations of our method. When there is a change point that is a dip, we cannot use topic modeling to explain it, as a dip in the emotion or moral sentiment indicates a decrease of discussion related to an event. However, usually the decrease of some emotions is accompanied by the increase of some other emotions, and we can study the tweets tagged with the surged emotions to understand the topics.

ACKNOWLEDGMENT

This work is supported in part by AFOSR under grants FA9550-22-1-0380 & FA9550-20-1-0224, and DARPA under contract HR001121C0168.

REFERENCES

- [1] M. Klačnja *et al.*, “Measuring Public Opinion with Social Media Data,” Oxford University Press, 09 2018.
- [2] M. Pellert *et al.*, “Validating daily social media macroscopes of emotions,” *Scientific Reports*, vol. 12, no. 1, p. 11236, 2022.
- [3] Dodds *et al.*, “Fame and ultrafame: Measuring and comparing daily levels of ‘being talked about’ for united states’ presidents, their rivals, god, countries, and k-pop,” *JQD:DM*, vol. 2, Feb. 2022.
- [4] M. Malik *et al.*, “A performance comparison of unsupervised techniques for event detection from oscar tweets,” *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [5] J. Weng *et al.*, “Event detection in twitter,” in *In ICWSM-2011*, vol. 5, no. 1, 2011, pp. 401–408.
- [6] K. Morabia *et al.*, “SEDTWik: Segmentation-based event detection from tweets using Wikipedia,” in *NACCL workshop*, 2019, pp. 77–85.
- [7] Z. Rezaei *et al.*, “Event detection in twitter by deep learning classification and multi label clustering virtual backbone formation,” *Evolutionary Intelligence*, pp. 1–15, 2022.
- [8] G. A. vanKleef *et al.*, “Editorial: The social nature of emotions,” *Frontiers in Psychology*, vol. 7, p. 896, 2016.
- [9] J. Haidt *et al.*, “The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules,” *The innate mind*, vol. 3, pp. 367–391, 2007.
- [10] M. B. Babac, “Emotion analysis of user reactions to online news,” *Information Discovery and Delivery*, no. ahead-of-print, 2022.
- [11] K. K. Aldous *et al.*, “Measuring 9 emotions of news posts from 8 news organizations across 4 social media platforms for 8 months,” *ACM Transactions on Social Computing (TSC)*, vol. 4, no. 4, pp. 1–31, 2022.
- [12] H. Alhuzali *et al.*, “SpanEmo: Casting multi-label emotion classification as span-prediction,” in *ECACL*. ACL, Apr. 2021, pp. 1573–1584.
- [13] S. Mohammad *et al.*, “SemEval-2018 task 1: Affect in tweets,” in *Proc. 12th Int. Workshop on Semantic Evaluation*, Jun. 2018, pp. 1–17.
- [14] S. Guo *et al.*, “A data fusion framework for multi-domain morality learning,” in *In ICWSM-2023*, vol. 17, 2023, pp. 281–291.
- [15] J. Hoover *et al.*, “Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment,” *Social Psychological and Personality Science*, vol. 11, no. 8, pp. 1057–1071, 2020.
- [16] J. Garten *et al.*, “Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis: Distributed dictionary representation,” *Behavior Research Methods, Instruments, and Computers*, vol. 50, no. 1, pp. 344–361, Feb. 2018.
- [17] D. V. Hinkley, “Inference about the change-point from cumulative sum tests,” *Biometrika*, vol. 58, no. 3, pp. 509–523, 1971.
- [18] R. P. Adams *et al.*, “Bayesian online changepoint detection,” *arXiv preprint arXiv:0710.3742*, 2007.
- [19] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based tf-idf procedure,” *arXiv preprint arXiv:2203.05794*, 2022.
- [20] A. B. Dieng *et al.*, “Topic modeling in embedding spaces,” *TACL*, vol. 8, pp. 439–453, 2020.
- [21] J. Leskovec *et al.*, “Meme-tracking and the dynamics of the news cycle,” in *Proceedings of the 15th ACM SIGKDD*, 2009, pp. 497–506.
- [22] N. Reimers *et al.*, “Sentence-bert: Sentence embeddings using siamese bert-networks,” in *In EMNLP-2019*. ACM, 11 2019.
- [23] C. Li *et al.*, “Twevent: Segment-based event detection from tweets,” ser. CIKM ’12. New York, NY, USA: ACM, 2012, p. 155–164.
- [24] J. Allan *et al.*, “Topic Detection and Tracking Pilot Study Final Report,” 1998.