

# Emotion-Aware Event Summarization in Microblogs

Rrubaa Panchendrarajan  
School of Computing, National  
University of Singapore  
rrubaa@comp.nus.edu.sg

Wynne Hsu  
School of Computing, National  
University of Singapore  
Institute of Data Science, National  
University of Singapore  
NUS Centre for Trusted Internet and  
Community  
whsu@comp.nus.edu.sg

Mong Li Lee  
School of Computing, National  
University of Singapore  
Institute of Data Science, National  
University of Singapore  
NUS Centre for Trusted Internet and  
Community  
leeml@comp.nus.edu.sg

## ABSTRACT

Microblogs have become the preferred means of communication for people to share information and feelings, especially for fast evolving events. Understanding the emotional reactions of people allows decision makers to formulate policies that are likely to be more well-received by the public and hence better accepted especially during policy implementation. However, uncovering the topics and emotions related to an event over time is a challenge due to the short and noisy nature of microblogs. This work proposes a weakly supervised learning approach to learn coherent topics and the corresponding emotional reactions as an event unfolds. We summarize the event by giving the representative microblogs and the emotion distributions associated with the topics over time. Experiments on multiple real-world event datasets demonstrate the effectiveness of the proposed approach over existing solutions.

## KEYWORDS

Topic-Emotion Learning, Microblogs, Event Summary

### ACM Reference Format:

Rrubaa Panchendrarajan, Wynne Hsu, and Mong Li Lee. 2021. Emotion-Aware Event Summarization in Microblogs. In *Companion Proceedings of the Web Conference 2021 (WWW '21 Companion)*, April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3442442.3452311>

## 1 INTRODUCTION

Microblogging platforms have become a popular means of communication among users especially for the sharing of news and opinions as events unfold in the real-world. This has led to extensive research in event analysis in microblogs which encompasses topic mining and event summarization. Existing works on summarizing an event based on the textual content of microblogs mostly provide a general factual summary [3, 12, 14, 19, 28] without considering users' emotional reactions. Yet, having a good understanding of the emotional responses of users is useful for policy makers.

Figure 1 shows an example of the summary with the corresponding users' reaction for the tweets collected over a period of 60 days

using keywords *corona*, *coronavirus* and *nCov* during the COVID-19 outbreak. We observe that at the start of the outbreak, the primary emotion was afraid as new confirmed cases and death have increased in China and subsequently found in Iran. As the death toll in China rose sharply, the dominant emotion was sad. This became afraid again when the first case was confirmed in New York, revealing that people recognized that COVID-19 might become a pandemic. When Italy reported a significant number of new covid cases on 16 March, people were mostly surprised, possibly at the high rate of disease infection. Their emotions turned to afraid when the healthcare system was reported to be on the verge of being overwhelmed. When countries started to go into lockdown, the dominant emotion was sad as people stayed at home. There was a rare moment of happy emotion when Lady Gaga announced the funds raised to help fight the pandemic. This example highlights the additional insights gained when we have an event summarization with both a comprehensive coverage of the topics as well as the corresponding users' reactions. Knowing the emotions of the people as an event unfolds could enable governments or organizations provide better support to them, especially during a crisis.

Generating an event summary with coherent topics and emotions in microblogs is a challenge as the tweets are short and noisy, with facts peppered with users' emotional reactions. Further, users often use the same emotional words across different topics, and the same words may be used to express both facts and emotions. For example, the tweets "*the children are safe*" and "*be safe*" contain the word "safe". However, the former conveys the fact that the children are safe, while the latter expresses the emotion hoping that one stays safe. Statistical topic models such as [20, 25, 31, 34, 38] learn topics based on the co-occurrences of words. Having words that convey both facts and emotions may lead to tweets from different topics being clustered together, thus affecting the quality of the models learned. This raises the need to differentiate whether a word is conveying fact or emotion.

In this work, we propose a weakly supervised topic-emotion model that takes into consideration words conveying factual information about a topic, and words expressing the emotions of the users. Our model does not rely on the availability of labeled training data to learn topics and emotions. We summarize the event using representative microblogs and emotion distributions of topics found. Extensive experiments on real world event datasets demonstrate that our approach is able to learn coherent topics as well as changes in emotional reactions as an event unfolds. Case studies on the Mexico Earthquake event shows that our summary is more

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

*WWW '21 Companion*, April 19–23, 2021, Ljubljana, Slovenia

© 2021 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-8313-4/21/04.

<https://doi.org/10.1145/3442442.3452311>

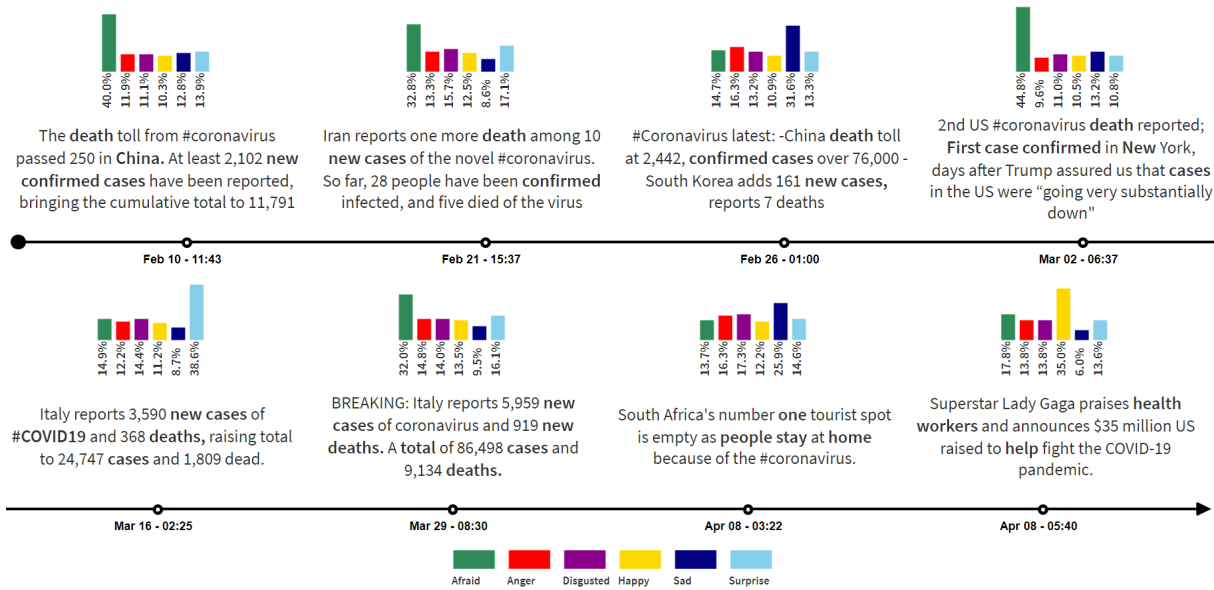


Figure 1: Summary of COVID-19 outbreak over 3.2M tweets collected from Feb 6th to Apr 9th 2020.

informative compared to state-of-the-art solutions. To the best of our knowledge, this is the first work on event summarization to include both facts and user emotions.

## 2 RELATED WORK

Event summarization in microblogs have mostly focused on generating a textual summary of the event [3, 4, 10, 12, 17, 32] and ignored user reactions. Traditional summarization approaches utilize either k-means clustering [12, 32] or detect volume peaks in microblogs [3, 4, 10] to identify subevents and pick the most representative microblogs to summarize the subevents.

Another approach employs topic models for clustering microblogs before picking representative tweets from each topic cluster [17]. However, the short and noisy nature of the microblogs make the learning of topics challenging. TwitterLDA [34] assumes each microblog is associated with a single topic and classifies the words in a microblog into background and topic words. [29] aggregates the microblogs with similar hashtags, while [11, 25] use word embedding [24] to create pseudo-documents and learn topics directly from these documents. [20, 38] use the topic model itself to generate the pseudo documents by modeling the pseudo document assignment of a short text as a latent variable. All these models do not consider user emotions when learning the topics.

Several works have tried to model both topics and emotion. [36] extracts topics using LDA [2], and the emotion of each microblog is determined using a supervised emotion classifier. The topic-emotion models in [9, 16, 26, 27, 35] learn a supervised topic-aware emotion classifier. These models rely on labeled training data and hence their ability to generalize to new topics is limited.

Other works have considered sentiments (positive, negative, and neutral) and adopted weakly supervised approach to learn coarse-grained topic-sentiment from documents. JST [21] extends LDA with a new sentiment layer where a word generation depends not

only on the topic but also on the sentiment. TS [5] is a variant of JST where the position of sentiment and topic layer is swapped. Both JST and TS are designed for documents and do not address the challenges in learning topics from short texts. [33] extends JST for microblogs and proposes a topic-sentiment model called LDST to jointly model topics, sentiment, time and geolocation information for tracking topics and sentiments held by users in different locations over time. [15] designs a model based on JST, called MJST, and incorporates emoticons as a separate modality. BST [13] imposes a bias layer on top of a topic-sentiment layer to classify words as either subjective or objective before assigning the topic-sentiment label. These methods do not differentiate words conveying the factual information about a topic from user sentiment.

## 3 EMOTION-TOPIC MODELING

We design a weakly supervised topic-emotion model, called ETM that takes into account words that convey factual information from words that convey emotions. Let  $M$  be the set of microblogs. Each microblog is pre-processed and transformed into a bag of words. Similar to Twitter-LDA [34], we assume that each microblog is associated with one topic. Let  $Z = \{z_1, z_2, \dots\}$  be the set of topics, and  $E = \{e_1, e_2, \dots\}$  be the set of emotions associated with the topics. Both  $z \in Z$  and  $e \in E$  are distributions over some vocabulary  $V = \{v_1, v_2, \dots\}$ . Each  $v \in V$  can be one of the following *type*:

- (1) *generic* if it appears in almost all the topics associated with some event, e.g. "election", "earthquake",
- (2) *specific* if it conveys some facts, e.g. "collapsed", or
- (3) *emotional* if it expresses some emotion, e.g. 😊, pray, ❤️.

This allows the same word to be used to convey fact or emotion or both. Moreover, the emotional words are independent of topics, so the same *emotional* word can be associated with different topics. During the learning, the model samples a word *type* for each word

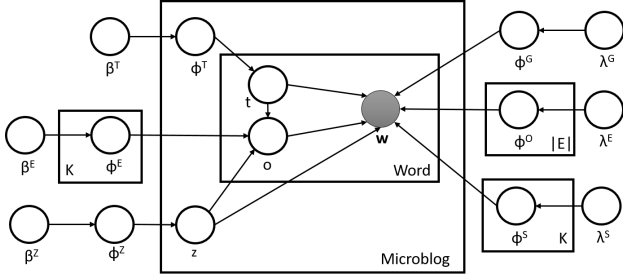


Figure 2: Graphical model representation of ETM

Table 1: List of key notations

Notation	Description
$Dir(x)$	Dirichlet distribution with prior $x$
$Multi(x)$	Multinomial distribution with prior $x$
$\phi^Z$	Topic distribution of an event
$\phi^G$	<i>generic</i> word distribution
$\phi^T$	Distribution of word type in a microblog belonging to <i>specific/generic/emotional</i>
$\phi_e^O$	<i>emotional</i> word distribution of an emotion $e$
$\phi_k^S$	<i>specific</i> word distribution of the topic $k$
$\phi_k^E$	Emotion distribution of a topic $k$
$K$	Number of topics
$c_{x,i,w_j}^{-(i,j)}$	No. of times $w_j$ is assigned to $x$ in $m_i$ excluding the position $(i, j)$
$c_{x,*,w_j}^{-(i,j)}$	No. of times $w_j$ is assigned to $x$ in all microblogs excluding the position $(i, j)$
$c_{x,i,*}^{-(i,j)}$	#words assigned to $x$ in document $m_i$ excluding the position $(i, j)$
$c_{x,*}^{-(i)}$	#microblogs assigned to $x$ excluding $m_i$

in a microblog as either *generic* or *specific* or *emotional*, discards the *generic* words and utilizes the *specific* words and emotion associated with the *emotional* words to determine the topic discussed in a microblog. Figure 2 gives the graphical model representation and the notations used are summarized in Table 1.

The generative process proceeds as follows:

- (1) Draw the topic distribution  $\phi^Z \sim Dir(\beta^Z)$
- (2) Draw the *generic* word distribution  $\phi^G \sim Dir(\lambda^G)$
- (3) For each emotion  $e \in \{1, 2, \dots\}$ ,
  - Draw the *emotional* word distribution  $\phi_e^O \sim Dir(\lambda^O)$
- (4) For each topic  $k \in \{1, 2, \dots, K\}$ ,
  - Draw the *specific* word distribution  $\phi_k^S \sim Dir(\lambda^S)$
  - Draw the emotion distribution  $\phi_k^E \sim Dir(\beta^E)$
- (5) For each microblog  $m_i$ ,
  - Draw the topic assignment  $z_i \sim Multi(\phi^Z)$
  - Draw the word-type distribution  $\phi_i^T \sim Dir(\beta^T)$
- (6) For each word  $w_j$  in  $m_i$ ,
  - Draw the word type assignment  $t_j \sim Multi(\phi_i^T)$
  - If  $t_j$  indicates *specific*, then draw a word  $w_j \sim Multi(\phi_k^S)$

- If  $t_j$  indicates *generic*, then draw a word  $w_j \sim Multi(\phi^G)$
- If  $t_j$  indicates *emotional*, then draw an emotion  $o_j \sim Multi(\phi_k^E)$  and then draw a word  $w_j \sim Multi(\phi_{o_j}^E)$

We employ Gibbs sampling to infer the latent variables  $z$ ,  $t$ , and  $o$  with the following update rules:

$$P(t_j = S|z, w, \dots) \propto (\beta_S^T + c_{S,i,*}^{-(i,j)}) \times \frac{\lambda_{w_j}^S + c_{z_i,*,w_j}^{-(i,j)}}{\sum_{v=1}^{|V|} \lambda_v^S + c_{z_i,*,v}^{-(i,j)}}$$

$$P(t_j = G|w, \dots) \propto (\beta_G^T + c_{G,i,*}^{-(i,j)}) \times \frac{\lambda_{w_j}^G + c_{G,*,w_j}^{-(i,j)}}{\sum_{v=1}^{|V|} \lambda_v^G + c_{G,*,v}^{-(i,j)}}$$

$$P(t_j = O|z, w, \dots) \propto (\beta_O^T + c_{O,i,*}^{-(i,j)}) \times \frac{\lambda_{w_j}^O + c_{o_j,*,w_j}^{-(i,j)}}{\sum_{v=1}^{|V|} \lambda_v^O + c_{o_j,*,v}^{-(i,j)}}$$

$$P(o_j = e|z, w, \dots) \propto (\beta_e^E + c_{z_i,*,e}^{-(i,j)}) \times \frac{\lambda_{w_j}^O + c_{e,*,w_j}^{-(i,j)}}{\sum_{v=1}^{|V|} \lambda_v^O + c_{e,*,v}^{-(i,j)}}$$

$$P(z_i = k|t, w, o, \dots) \propto (\beta_k^Z + c_{k,*}^{-(i)}) \times \prod_{t_j=S} \frac{\lambda_{w_j}^S + c_{k,*,w_j}^{-(i,j)}}{\sum_{v=1}^{|V|} \lambda_v^S + c_{k,*,v}^{-(i,j)}} \times \prod_{t_j=O} \frac{\lambda_{o_j}^E + c_{k,*,o_j}^{-(i,j)}}{\sum_{e=1}^{|E|} \lambda_e^E + c_{k,*,e}^{-(i,j)}}$$

Once the latent variables are learned using Gibbs sampling, the *specific* word distribution and the emotion distribution of  $k^{th}$  topic can be computed as follows:

$$\phi_k^S(v_j) = \frac{\lambda_{v_j}^S + c_{k,*,v_j}}{\sum_{v=1}^{|V|} \lambda_v^S + c_{k,*,v}} \quad \phi_k^E(e_j) = \frac{\beta_{e_j}^E + c_{k,*,e_j}}{\sum_{e=1}^{|E|} \beta_e^E + c_{z_i,*,e}}$$

where  $\phi_k^S(v_j)$  and  $\phi_k^E(e_j)$  are the probabilities of the word  $v_j$  and emotion  $e_j$  appearing in the *specific* word and emotion distribution of the  $k^{th}$  topic respectively.

#### 4 EVENT SUMMARY GENERATION

Temporal peaks in microblogs is a good indicator of the emergence of subevents. We adapt the peak detection algorithm OPAD [23] to determine the subevents.

We first apply our topic-emotion model ETM in each subevent to discover the hot topics associated with it. After learning the topics and its corresponding emotional reactions in each subevent, we generate the event summary over time by summarizing each topic with a representative microblog and its emotion distribution learned. We compute the similarity between a microblog  $m_i$  and the *specific* word distribution  $\phi_k^S$  of the topic  $k$  as follows:

$$similarity(m_j, \phi_k^S) = (1 - D_{KL}(\Theta_{m_j} || \phi_k^S))_{norm} \quad (1)$$

where  $\Theta_{m_j}$  is the word distribution of  $m_j$ , and  $D_{KL}()$ <sub>norm</sub> is the normalized value of Kullback-Leibler (KL) divergence [18], which measures the distance between two distributions  $D_1$  and  $D_2$ :

$$D_{KL}(D_1||D_2) = \sum_w p(w|D_1) \log \frac{p(w|D_1)}{p(w|D_2)} \quad (2)$$

A lower KL divergence score indicates that the two distributions  $D_1$  and  $D_2$  are more similar to each other.

The microblog with the highest similarity score is selected as the representative microblog. If there are more than one microblogs with the same highest scores, we pick the one that is posted by a user with the largest number of followers and reposts.

## 5 EVALUATION

In this section, we evaluate the effectiveness of our proposed approach using the datasets published by [1, 30, 37]. We select different types of events that occurred in the last 5 years. We use the TwitterAPI<sup>1</sup> to retrieve the tweets based on the published tweet IDs and retain only those tweets that are in English. We preprocess the tweets by removing user mention, URL, stop words and words with frequency less than 5. The remaining words in the tweets, hashtags (without their # symbols) and emojis form the vocabulary. Table 2 gives the characteristics of these event datasets.

**Table 2: Statistics of the event datasets**

Event dataset	# Tweets	Period	# Subevent
Wimbledon	668,280	11-12 Jul 2015	3
Oscar Celebration	271,508	28 Feb 2016	2
Irish Election	566,195	3 Feb to 6 Mar 2016	4
Ontario Election	1,071,314	28 Apr to 14 Jun 2018	4
Mexico Earthquake	263,354	30 Sep to 6 Oct 2017	3
COVID-19 Outbreak	3,182,017	4 Feb to 9 Apr 2020	8

### 5.1 Experiments on Topic Models

We compare our emotion-topic model **ETM** with the following topic-sentiment methods:

- **JST** [21]. This model extends LDA with a sentiment layer where the word generated depends on topics and sentiments.
- **TS** [5]. This is a variant of JST where the position of the topic and sentiment layers are swapped.
- **LDST** [33]. This sentiment topic model jointly models topics, sentiments, time and geolocation information.
- **BST** [13]. This model extends JST with an additional bias layer where a word is tagged as either subjective or objective.
- **MJST** [15]. This model extends JST to incorporate emoticons.

These topic-sentiment models samples a topic and, a sentiment for each word in a microblog during the learning and generate word distribution per each topic-sentiment pair and, sentiment distribution per topic at the end of the learning. We adapt these model by replacing the sentiment layer with an emotion layer.

Even though our model does not require any labeled training data to learn topics and emotions, we provide a few seed words to the model during the initialization of latent variables. We use 5

emojis per emotion from the lexicon [7] and 40 words per emotion as the *emotional* seed words. By keeping Ekman’s big six emotions [6], *afraid, anger, disgusted, happy, sad and surprise* as the initial seed word for each emotion respectively, emotional seed words are expanded by iteratively adding their synonyms. For fair comparison, we provide the same emotional seed words to all the models during the initialization of latent variables. We also extract the top 10 frequent hashtags as seed words for the *generic* words. Table 3 shows sample of the emotional seed words.

**Table 3: Sample emotional seed words**

	Sample words
Afraid	fearful, helpless, terrible, 😨, 😱
Anger	angry, annoy, hate, 😡, 😠
Disgusted	disgust, shame, embarrass, 😒
Happy	joy, cheer, love, 😊, 😄
Sad	sorrow, sorrowful, hurt, 😞, 😓
Surprise	amaze, stunned, magnificent, ***, 😲

A manual analysis reveals that the majority of the subevents have 4-5 topics. The settings for the 6 concentration parameters are:  $\beta^Z = 1$ ,  $\beta^E = 0.1$ ,  $\beta^T = 0.1$ ,  $\lambda^G = 0.1$ ,  $\lambda^S = 0.01$ , and  $\lambda^E = 0.01$ . A small value indicates that the distribution is concentrated on a small set of values [17]. The number of burn-in and total iterations for Gibbs sampling are 50 and 300 respectively as done in [5, 27, 35]. The results reported are averaged over five runs.

**5.1.1 Evaluation of Topics Learned.** We first examine the effectiveness of the various methods to learn coherent and comprehensive topics for an event. Besides the topic-emotion models, we also compare ETM with the following:

- **Twitter-LDA** [34]. This method assumes a single topic assignment for an entire tweet.
- **LDA-Agg** [29]. This method uses hashtags to cluster the tweets and then learns topics using LDA [2].
- **EMB** [25]. This method uses embeddings to cluster short texts into a long pseudo-text before inferring topics.

We use the Pointwise Mutual Information metric (PMI) [8, 11, 38] to measure the coherence of topics learned by the various methods. The intuition behind PMI is that if the top words of a topic co-occur frequently relative to the number of times the words appear individually, then the words are more likely to form a cohesive topic. Let  $W = \{w_1, \dots, w_N\}$  be the top- $N$  words of a topic  $z$ ,  $P(w)$  be the probability of word  $w$  determined using the number of occurrence of the word  $w$ , and  $P(w_1, w_2)$  be the probability of words  $w_1$  and  $w_2$  occurring together. Then the PMI of  $z$  is given by

$$PMI(W) = \frac{2}{N * (N - 1)} \sum_{i=2}^N \sum_{j=1}^{i-1} \log \frac{P(w_i, w_j)}{P(w_i) * P(w_j)} \quad (3)$$

Table 4 shows the PMI of the various methods when number of topics  $K$  is set to 5. We see that ETM significantly outperforms existing topic models as well as topic-sentiment models. Note that LDA-Agg model has no result as it is unable to handle the large number of tweets in the COVID-19 dataset.

<sup>1</sup><https://developer.twitter.com/en/docs/api-reference-index.html>

**Table 4: PMI of top N specific words**

	Top-N Words	Topic Models			Adapted Sentiment Topic Models					
		Twitter-LDA	LDA-Agg	EMB	JST	TS	LDST	BST	MJST	ETM
Wimbledon	5	0.8126	0.4091	0.7714	0.3869	0.7293	0.345	0.3438	0.4704	<b>0.9615</b>
	10	0.465	0.2792	0.5494	0.1795	0.5536	-0.1567	0.0707	0.1803	<b>0.7333</b>
	20	0.0797	0.0273	0.1589	-0.1681	0.2342	-0.4632	-0.2766	-0.2114	<b>0.3859</b>
Oscar Celebration	5	0.7255	0.6714	0.6718	0.3515	0.6198	0.2001	0.4744	0.3531	<b>0.9903</b>
	10	0.4904	0.3142	0.378	0.0973	0.3089	0.1815	0.2119	0.0012	<b>0.7285</b>
	20	0.1825	0.1792	0.1333	-0.1606	0.1255	-0.1092	-0.0763	-0.1906	<b>0.28</b>
Irish Election	5	0.6811	0.4406	0.4303	0.2072	0.5384	0.2646	0.0093	0.2297	<b>0.9905</b>
	10	0.4728	0.3023	0.4739	0.148	0.4347	0.0183	0.0531	0.1231	<b>0.7094</b>
	20	0.2999	0.2199	0.2807	0.0092	0.2627	-0.0783	-0.0611	0.0105	<b>0.4402</b>
Ontario Election	5	0.4314	0.424	0.4851	0.4347	0.4454	0.2533	0.3683	0.4821	<b>0.7764</b>
	10	0.3546	0.3633	0.3865	0.2862	0.3216	0.1664	0.2198	0.3041	<b>0.657</b>
	20	0.2416	0.2667	0.3013	0.1513	0.227	0.0674	0.0845	0.1716	<b>0.4621</b>
Mexico Earthquake	5	0.8849	0.6807	0.7584	0.3832	0.7473	0.478	0.2412	0.2128	<b>1.367</b>
	10	0.6005	0.4935	0.5261	0.0471	0.4489	-0.0749	-0.1233	-0.044	<b>1.0337</b>
	20	0.2009	0.3193	0.2134	-0.3156	0.175	-0.4231	-0.4118	-0.3188	<b>0.4326</b>
COVID-19 Outbreak	5	0.5838	–	0.4748	0.144	0.4318	-0.0043	0.01	0.1464	<b>0.8565</b>
	10	0.4874	–	0.4187	0.1145	0.3707	0.0362	-0.014	0.0548	<b>0.6645</b>
	20	0.385	–	0.3574	0.0884	0.3145	-0.0184	-0.0395	0.0207	<b>0.477</b>

**Table 5: Top 10 emotional words learned in the Mexico earthquake dataset**

	ETM	TS	MJST
Afraid	<b>terrible</b> , fresh, <b>panic</b> , small, 🏠, much, hit, government, 🇺🇸, handling	city, news, new, survivor, mexi-coearthquake, victim, week, help, hit, aid	rescue, survivor, search, mexican, hit, help, effort, killed, victim, quake
Anger	city, <b>anger</b> , <b>aftershock</b> , building, response, rubble, third, man, nature, rescued	oaxaca, people, search, city, magnitude, dog, help, frida, new, jittery	victim, help, frida, relief, never, sofia, survivor, rescuer, donate, response
Disgusted	bring, frida, many, consolation, mother, sorry, 🙄, 🙄, 2, team, job	new, magnitude, 2017, city, shake, time, relief, mexicoearthquake, toll, help	dog, rescue, strike, building, frida, life, rubble, day, ago, km
Happy	❤️, 🎉, 🏠, <b>love</b> , 🙄, 🙄, 🙄, <b>great</b> , 🙄, <b>thankful</b>	magnitude, victim, city, rescue, hit, dog, new, strike, effort, min, 19	help, people, hurricane, affected, de, puerto, please, rico, victim, relief
Sad	❤️, 🙄, 🙄, 🙄, <b>praying</b> , <b>lost</b> , 🏠, make, <b>sad</b> , 🙄	shake, new, city, survivor, help, rescue, quake, death, 2, ago	magnitude, new, shake, jittery, hit, 2, strong, quake, relief, southern
Surprise	political, <b>good</b> , prayer, <b>rock</b> , <b>thank</b> , girl, <b>damn</b> , everyone, <b>omg</b> , <b>beautiful</b>	rescue, victim, quake, city, people, latest, help, death, hurricane, shake	magnitude, 2017, oaxaca, september, strike, 2, coast, new, 09, hit

5.1.2 *Evaluation of Emotions Learned.* Next, we examine the emotion distributions of topics learned by the various models. Table 5 compares the *emotional* words learned by ETM, TS and MJST. We observe that the majority of the *emotional* words learned by ETM convey user emotions. In contrast, the *emotional* words learned by TS and MJST are dominated by factual words and do not convey the emotion listed.

We also examine the diversity of emotion distributions  $\phi^E$  of  $K$  topics learned. This is given by:

$$Diversity(\phi^E) = \frac{1}{K * (K - 1)} \sum_{i=1}^K \sum_{j=1, i \neq j}^K D_{KL}(\phi_i^E || \phi_j^E) \quad (4)$$

where  $D_{KL}$  is computed using Equation 2.

A diversity value close to zero indicates the distributions are similar to each other. Figure 3 shows the performance of the models as we vary the number of seed words per emotion from 10 to 45. We observe that ETM has the highest diversity compared to the

rest. In particular, Mexico Earthquake has high diversity values [0.5-0.7] implying that users have expressed very different emotions for different topics, e.g., sad for 'death toll', happy for 'rescue with Frida'. The diversity values for both the election datasets are low [0.2-0.3], indicating that the range of emotions displayed across multiple topics are similar. On the other hand, the diversity values for the other methods are almost close to zero with similar emotional distributions across multiple topics.

The reason why ETM is able to identify the diverse emotional reactions is because we consider the type of a word, whether it is *generic*, *specific* or *emotional* while learning the emotional reactions associated with a topic. Only *emotional* words are assigned an emotion, ensuring that words conveying user emotion are accounted for in the emotion distribution.

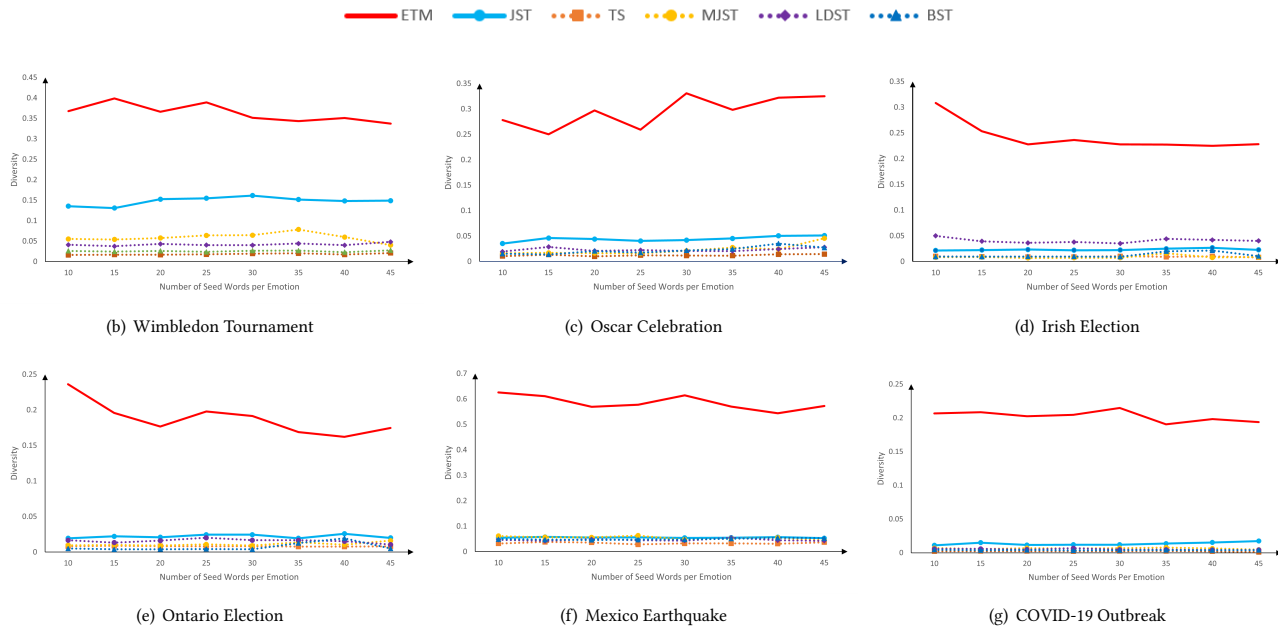


Figure 3: Diversity for varying number of emotional seed size

## 5.2 Experiments on Event Summarization

Finally, we examine the summaries generated by various models using the Mexico Earthquake dataset. Here, the number of topics per interval is 4. As there is no ground truth topics nor emotions for the datasets and it is not feasible to manually go through thousands of tweets to identify hot topics and the corresponding dominant emotions, we select the smallest dataset, the Mexico Earthquake dataset for this experiment. Three volunteers individually are asked to select four representative tweets and identify the dominant emotion expressed towards the topic, if any.

We use the ROUGE metric [22] for evaluating the quality of the textual content of the summaries generated by various models. Table 6 shows four variants of ROUGE for the summaries generated by ETM and other topic and topic-sentiment models. We see that ETM has attained the highest scores for all the variants.

Table 6: Average ROUGE scores for Mexico earthquake.

	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-3
Twitter-LDA	0.3633	0.4324	0.1785	0.1183
LDA-Agg	0.3365	0.4226	0.1597	0.0923
EMB	0.3051	0.4357	0.1447	0.0467
JST	0.3642	0.4462	0.1894	0.1178
TS	0.3538	0.4391	0.1887	0.1231
LDST	0.333	0.4183	0.1517	0.0788
BST	0.358	0.4418	0.1774	0.0946
MJST	0.3461	0.4379	0.1706	0.0897
ETM	<b>0.4241</b>	<b>0.4787</b>	<b>0.2391</b>	<b>0.1652</b>

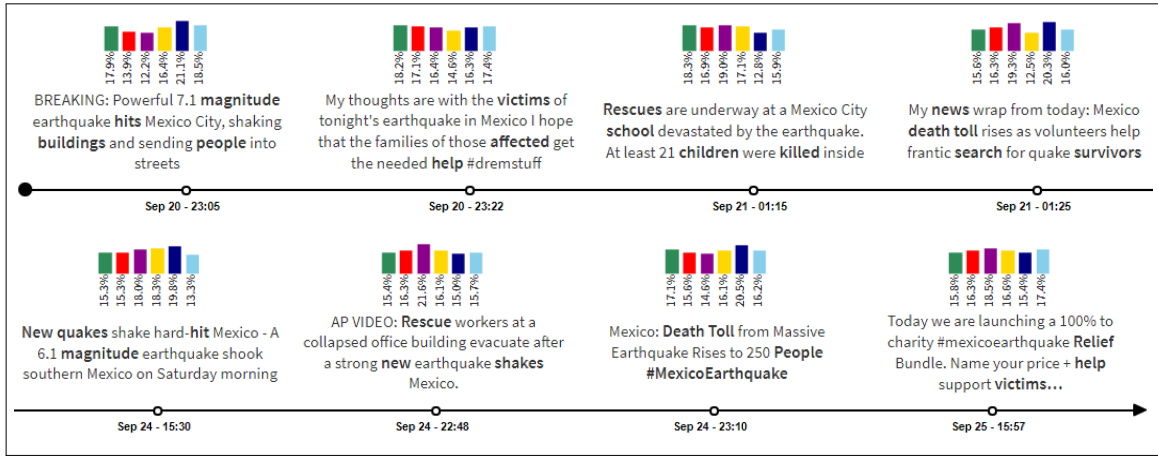
We compare the summaries generated by ETM with TS and MJST in the Mexico Earthquake dataset. Figure 4 shows the summaries generated by TS, MJST, and ETM in the first two subevents. The

Table 7: Topics covered in the first two subevents for Mexico earthquake.

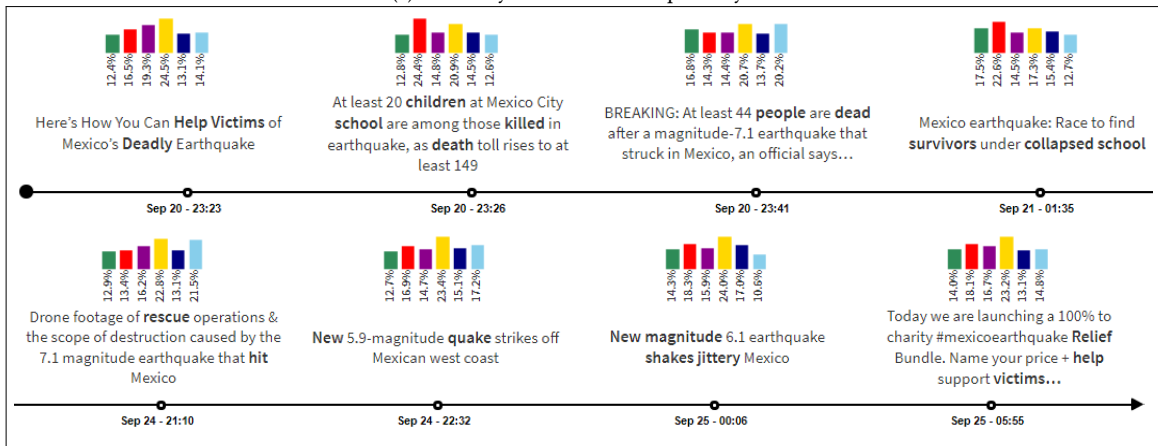
Subevent	Ground Truth Topics	TS	MJST	ETM
1	Help victim	✓	✓	✓
	School collapsed	✓	✓	✓
	Search for survivors	✓	✓	✓
	Death toll	✓	✓	✓
2	Help victim	✓	✓	✓
	Death toll	✓	✗	✓
	New Earthquake	✓	✓	✓
	Rescue with Frida	✗	✓	✓
	Coverage	0.875	0.875	<b>1</b>

top 5 words of each topic are highlighted in the representative microblog. We observe that for TS and MJST, the topics learned are a mixture of the different subject matters, and the emotions learned are almost evenly distributed with no clear dominant emotion, which is not very useful for decision makers. One possible reason for the indistinguishable emotional responses is because it does not distinguish words that convey factual information from user emotion. Hence, words conveying factual information getting an emotion assignment result in evenly distributed emotion distributions. On the other hand, the summary generated by ETM captures the dominant emotion for most of the ground truth topics and the change in emotion across the topics over time, e.g., afraid for 'search for survivors' changes to happiness for 'rescue with Frida'.

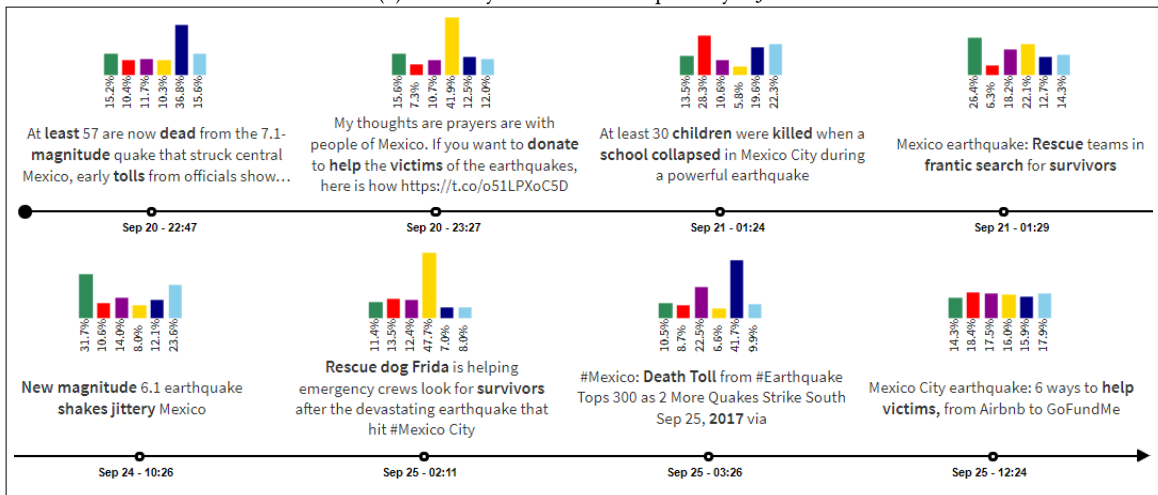
Tables 7 and 8 show the topics and corresponding dominant emotions identified by ETM, TS and MJST in the first two subevents of the Mexico Earthquake against the ground truth topics and dominant emotions obtained from human volunteers. We see that the summary generated by our proposed ETM has the highest coverage



(a) Summary of Mexico Earthquake by TS



(b) Summary of Mexico Earthquake by MJST



(c) Summary of Mexico Earthquake by ETM

Figure 4: Summaries generated by TS, MJST and ETM.

**Table 8: Dominant emotions in the first two subevents for Mexico earthquake.**

Subevent	Dominant Emotion	TS	MJST	ETM
1	Happy	✗	✓	✓
	Sad	✗	✗	✗
	Afraid	✗	✗	✓
	Sad	✓	✗	✓
2	Happy	✗	✓	✗
	Sad	✓	✗	✓
	Afraid, Sad	✓	✗	✓
	Happy	✗	✗	✓
	Accuracy	0.375	0.25	<b>0.75</b>

of topics and is able to better capture the emotional reactions of the users with an accuracy of 0.75.

## 6 CONCLUSION

In this work, we have described an emotion-topic model called ETM to summarize the topics and emotions of an event. The proposed model has taken into consideration whether a word is generic, specific or expresses the emotion of a user. A weakly supervised learning method has been designed to capture the coherent topics and emotional reactions as an event unfolds. Extensive experiments on multiple real world datasets have demonstrated that our emotion-topic model outperforms existing topic and topic-sentiment models in terms of PMI and KL Divergence metrics. We have also evaluated the quality of the generated summaries against the ground truth using ROUGE scores and showed that the event summary provided by ETM gives better coverage of the topics and conveys more diverse user emotions. Future work includes multi-media and multi-platform event summarization, as well as exploring methods to generate a new summary instead of selecting representative tweets.

## REFERENCES

- [1] Firoz Alam, Ferda Ofli, and Muhammad Imran. 2018. Crisismm: Multimodal twitter datasets from natural disasters. In *Twelfth International AAAI Conference on Web and Social Media*.
- [2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022.
- [3] Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, and Mimmo Parente. 2016. Time aware knowledge extraction for microblog summarization on twitter. *Information Fusion* 28 (2016), 60–74.
- [4] Wenjie Li Dehong Gao and Renxian Zhang. 2013. Sequential summarization: A new application for timely updated twitter trending topics. In *Proceedings of the 51st annual meeting of the Association for Computational Linguistics*. 567–571.
- [5] Mohamed Dermouche, Leila Kouas, Julien Velcin, and Sabine Loudcher. 2015. A joint model for topic-sentiment modeling from text. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*. ACM, 819–824.
- [6] Paul Ekman, E Richard Sorenson, and Wallace V Friesen. 1969. Pan-cultural elements in facial displays of emotion. *Science* 164, 3875 (1969), 86–88.
- [7] Abdallah El Ali, Torben Wallbaum, Merlin Wasmann, Wilko Heuten, and Susanne CJ Boll. 2017. Face2Emoji: Using Facial Emotional Expressions to Filter Emojis. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI EA '17*). ACM, New York, NY, USA, 1577–1584. <https://doi.org/10.1145/3027063.3053086>
- [8] Anjie Fang, Craig Macdonald, Iadh Ounis, and Philip Habel. 2016. Topics in tweets: A user study of topic coherence metrics for Twitter data. In *European Conference on Information Retrieval*. Springer, 492–504.
- [9] Jiachun Feng, Yanghui Rao, Haoran Xie, Fu Lee Wang, and Qing Li. 2019. User group based emotion detection and topic discovery over short text. *World Wide Web* (2019), 1–35.
- [10] D. Gao, W. Li, X. Cai, R. Zhang, and Y. Ouyang. 2014. Sequential Summarization: A Full View of Twitter Trending Topics. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22, 2 (Feb 2014), 293–302.
- [11] Wang Gao, Min Peng, Hua Wang, Yanchun Zhang, Qianqian Xie, and Gang Tian. 2018. Incorporating word embeddings into topic modeling of short text. *Knowledge and Information Systems* (18 Dec 2018). <https://doi.org/10.1007/s10115-018-1314-7>
- [12] Mehreen Gillani, Muhammad U. Ilyas, Saad Saleh, Jalal S. Alowibdi, Naif Aljohani, and Fahad S. Alotaibi. 2017. Post Summarization of Microblogs of Sporting Events. In *Proceedings of the 26th International Conference on World Wide Web Companion* (Perth, Australia) (*WWW '17 Companion*). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 59–68.
- [13] Juncai Guo and Xue Chen. 2018. Bias-Sentiment-Topic model for microblog sentiment analysis. *Concurrency and Computation: Practice and Experience* 30, 13 (2018), e4417.
- [14] Ruifang He and Xingyi Duan. 2018. Twitter summarization based on social network and sparse reconstruction. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [15] Faliang Huang, Shichao Zhang, Jilian Zhang, and Ge Yu. 2017. Multimodal learning for topic sentiment analysis in microblogging. *Neurocomputing* 253 (2017), 144–153.
- [16] Minghui Huang, Yanghui Rao, Yuwei Liu, Haoran Xie, and Fu Lee Wang. 2018. Siamese network-based supervised topic modeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 4652–4662.
- [17] Hanwang Zhang Jingwen Bian, Yang Yang and Tat-Seng Chua. 2015. Multimedia Summarization for Social Events in Microblog Stream. In *IEEE Transactions on Multimedia*, Vol. 17. 216 – 228.
- [18] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.
- [19] Quanzhi Li and Qiong Zhang. 2020. Abstractive Event Summarization on Twitter. In *Companion Proceedings of the Web Conference 2020*. 22–23.
- [20] Ximing Li, Changchun Li, Jinjin Chi, and Jihong Ouyang. 2018. Short text topic modeling by exploring original documents. *Knowledge and Information Systems* 56, 2 (2018), 443–462.
- [21] Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 375–384.
- [22] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*. 74–81.
- [23] Adam Marcus, Michael S Bernstein, Osama Badar, David R Karger, Samuel Madden, and Robert C Miller. 2011. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 227–236.
- [24] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Atlanta, Georgia). Association for Computational Linguistics, 746–751.
- [25] Jipeng Qiang, Ping Chen, Tong Wang, and Xindong Wu. 2017. Topic Modeling over Short Texts by Incorporating Word Embeddings. In *Advances in Knowledge Discovery and Data Mining*, Jinho Kim, Kyuseok Shim, Longbing Cao, Jae-Gil Lee, Xuemin Lin, and Yang-Sae Moon (Eds.). Springer International Publishing, Cham, 363–374.
- [26] Yanghui Rao, Jianhui Pang, Haoran Xie, An Liu, Tak-Lam Wong, Qing Li, and Fu Lee Wang. 2017. Supervised Intensive Topic Models for Emotion Detection over Short Text. In *DASFAA*.
- [27] Yanghui Rao, Haoran Xie, Jun Li, Fengmei Jin, Fu Lee Wang, and Qing Li. 2016. Social emotion classification of short text via topic-level maximum entropy model. *Information & Management* 53, 8 (2016), 978–986.
- [28] Koustav Rudra, Subham Ghosh, Niloy Ganguly, Pawan Goyal, and Saptarshi Ghosh. 2015. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 583–592.
- [29] Asbjørn Steinskog, Jonas Therkelsen, and Björn Gambäck. 2017. Twitter Topic Modeling by Tweet Aggregation. In *Proceedings of the 21st Nordic Conference on Computational Linguistics* (Gothenburg, Sweden). Association for Computational Linguistics, 77–86.
- [30] Akanksha Tiwari, Christian Von Der Weth, and Mohan S. Kankanhalli. 2018. Multimodal Multiplatform Social Media Event Summarization. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 2s, Article 38 (April 2018), 23 pages. <https://doi.org/10.1145/3115433>
- [31] Yuan Wang, Jie Liu, Jishi Qu, Yalou Huang, Jimeng Chen, and Xia Feng. 2014. Hashtag Graph Based Topic Model for Tweet Mining. *2014 IEEE International Conference on Data Mining* (2014), 1025–1030.
- [32] Z. Wang, L. Shou, K. Chen, G. Chen, and S. Mehrotra. 2015. On Summarization and Timeline Generation for Evolutionary Tweet Streams. *IEEE Transactions on Knowledge and Data Engineering* 27, 5 (May 2015), 1301–1315. <https://doi.org/10.1109/TKDE.2014.2345379>



- [33] Min Yang, Jincheng Mei, Heng Ji, Zhou Zhao, Xiaojun Chen, et al. 2017. Identifying and Tracking Sentiments and Topics from Social Media Texts during Natural Disasters. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 527–533.
- [34] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing Twitter and Traditional Media Using Topic Models. In *Advances in Information Retrieval*. Springer Berlin Heidelberg, Berlin, Heidelberg, 338–349.
- [35] Wenjie Zheng, Zenan Xu, Yanghui Rao, Haoran Xie, Fu Lee Wang, and Reggie Kwan. 2017. Sentiment classification of short text using sentimental context. *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)* (2017), 1–6.
- [36] Qingqing Zhou and Chengzhi Zhang. 2017. Emotion evolutions of sub-topics about popular events on microblogs. *The Electronic Library* 35 (07 2017), 00–00. <https://doi.org/10.1108/EL-09-2016-0184>
- [37] Arkaitz Zubiaga. 2018. A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology* 69, 8 (2018), 974–984.
- [38] Yuan Zuo, Junjie Wu, Hui Zhang, Hao Lin, Fei Wang, Ke Xu, and Hui Xiong. 2016. Topic Modeling of Short Texts: A Pseudo-Document View. In *KDD*.