EnDSUM: Entropy and Diversity based Disaster Tweet Summarization

Piyush Kumar Garg[1], Roshni Chakraborty[2] and Sourav Kumar Dandapat[3]

[1]*Indian Institute of Technology, Patna, Bihar, India*
[2]*Aalborg University, Denmark*
[3]*Indian Institute of Technology, Patna, Bihar, India*

Abstract
The huge amount of information shared in Twitter during disaster events are utilized by government agencies and humanitarian organizations to ensure quick crisis response and provide situational updates. However, the huge number of tweets posted makes manual identification of the relevant tweets impossible. To address the information overload, there is a need to automatically generate summary of all the tweets which can highlight the important aspects of the disaster. In this paper, we propose an entropy and diversity based summarizer, termed as *EnDSUM*, specifically for disaster tweet summarization. Our comprehensive analysis on 6 datasets indicates the effectiveness of *EnDSUM* and additionally, highlights the scope of improvement of *EnDSUM*.

Keywords
Entropy, Disaster tweets, Social media, Summarization

1. Introduction

Social media platforms, like Twitter, are highly important mediums of information during disasters. For example, humanitarian organizations and government agencies rely on Twitter to identify relevant information on different categories, such as affected population, urgent need of resources, infrastructure damage, etc [1]. However, the huge number of tweets posted and the high vocabulary diversity [2, 3] makes it challenging to manually find the relevant information [4, 5]. In order to address this issue, several research works [6, 7] have proposed specific tweet summarization approaches for disaster events.

Existing disaster tweet summarization approaches could be segregated into content based [6], graph based [7], deep learning based [8], and ontology based [9] approaches on the basis of the mechanism they follow. While content based approaches [6, 10] rely only on the importance of the words present in a tweet to determine its selection to the summary, deep learning based approaches [8] consider both content and contextual importance of the tweet. However, none of these approaches consider the vocabulary diversity and therefore, fails to always ensure diversity in summary and coverage of all the important categories present in the tweets. In order

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'22 Workshop, Stavanger (Norway), 10-April-2022
piyush1_2021cs05@iitp.ac.in (P. K. Garg); roshnic@cs.aau.dk (R. Chakraborty); sourav@iitp.ac.in (S. K. Dandapat)
0000-0003-2266-9605 (P. K. Garg); 0000-0003-2043-2356 (S. K. Dandapat)

to address these, graph based approaches [7, 11] initially group similar tweets together such that each group represents a category by community detection algorithms, thereby handling the vocabulary diversity followed by selecting representative tweets from each group to create the summary to ensure coverage. However, automatic community detection algorithms fails to automatically segregate the tweets into different categories due to the vocabulary overlap among tweets of different categories. Therefore, Garg et al. [9] initially identify the category of each tweet by an ontology based approach and then, select tweets from each category to generate the summary. However, none of these approaches try to handle the vocabulary diversity simultaneously while selecting the tweets into the summary. For example, these existing approaches are dependent on identifying the categories initially which lead to bad summaries, such as reduced diversity in summary, if the categories are not identified correctly.

In order to resolve this, we propose *EnDSUM*, an entropy and diversity based disaster summarizer where we automatically select that tweet into summary which provides the best information coverage of all the tweets, i.e., *entropy* and most novel information, i.e., *diversity*. Therefore, *EnDSUM* can generate the summary automatically without explicitly identifying the category of a tweet. Although there are few single and multiple document summarization approaches [12, 13, 14, 15, 16] that have highlighted the relevance of *entropy* based selection of sentences into summary, those approaches are not directly applicable to disaster tweets. The reason being the informal structure of tweets, absence of storyline in tweets and the high vocabulary diversity in user generated tweets. Our evaluation of *EnDSUM* with existing state-of-the-art disaster tweet summarization approaches on 6 different disasters shows its high effectiveness on 5 datasets. However, we observe that the performance of *EnDSUM* degrades when there is considerable vocabulary overlap among the tweets which belong to different categories of the same disaster event. The reason being we consider only content based information for calculation of *entropy* and *diversity*. The organization of the paper is as follows. We discuss problem definition and proposed approach in Section 2 followed by the experiment details in Section 3 and conclusions in Section 4.

2. Proposed Approach

Given a disaster event, E, that consists of m tweets, $T = \{T_1, T_2, ..., T_m\}$, we aim to prepare a summary, S, by selecting l tweets from T such that it provides the maximum information coverage from T with minimum redundant information in the final summary. Therefore, we propose *EnDSUM* where we iteratively select the tweet that can ensure maximum entropy of all the tweets and maximum diversity in summary. While selection of the tweet with maximum *entropy* ensures information coverage of a category, selection of the tweet with the maximum *diversity* ensures not multiple tweets from the same category are selected [17, 18]. Although [17, 18] ensure maximization of *diversity* in summary, they propose network stratification based approaches which require explicit grouping of similar tweets together by community detection to ensure maximum *diversity*. [17, 18] are tweet summarization approaches related to news events which are not directly applicable to disaster events as community detection algorithms fail to group similar tweets in a disaster automatically [9]. Therefore, in *EnDSUM*, we propose an *entropy* and *diversity* based selection mechanism specific for tweets related to disaster events that does not require identification of similar groups and ensure better summary quality. Therefore, at every iteration, we select the tweet (T^*), which has the maximum score by Equation 1.

$$T^* = \arg\max \sum_{T_i \in (T - S')} (\alpha * E(T_i, K) + \beta * D(T_i, S^{'})) \tag{1}$$

where, $E(T_i, K)$ represents the *entropy* of T_i and K is the list of similar tweets of T_i, where a tweet is said to be similar to T_i if the content based cosine similarity, i.e. P_{ij} between them is higher than 0 (as shown in [19]) and P_{ij} is the normalized number of overlapping between T_i and T_j normalized by the total number of overlapping keywords of T_i with any tweet. $D(T_i, S^{'})$ represents the information *diversity* provided by T_i with respect to the already selected tweets in summary, $S^{'}$. α and β are the tunable parameters which represent the importance of $E(T_i, K)$ and $D(T_i, S^{'})$ respectively. We consider α and β as 0.5 to provide equal importance to both *entropy* and *diversity*. Although there are several available mechanisms to calculate $E(T_i)$, we rely on *Karci Entropy* [19] for *EnDSUM*. *Karci Entropy* can resolve the inherent vocabulary diversity in disaster tweets as it calculates the *entropy* of a tweet, $E(T_i, K)$, by considering the similarity of T_i with the other tweets as shown in Equation 2.

$$E(T_i, K) = \sum_{j=1}^{|K|} | - P_{ij}^{\gamma} \log P_{ij}|, \; 0 < \gamma \tag{2}$$

where, γ represents the importance of similarity. We consider γ as 0.5 as highlighted by Hark et al. [19]. Hark et al. [19] discuss while a lower value of γ mostly considers the impact of the local effect of the keywords, a higher value considers the impact of the global effect. Furthermore, they observe that the Rouge-N score was maximum for the γ value of 0.5 irrespective of the summary length which we directly consider as the γ value in *EnDSUM*. As a future direction of *EnDSUM*, we intend to exhaustively experiment and develop *Karci Entropy* such that it is most suitable for tweets related to disaster summarization. We calculate $D(T_i, S^{'})$ as (1-$Sim(T_i, S^{'})$) where $Sim(T_i, S^{'})$ represents the overlap in keywords between T_i and $S^{'}$ by

$$Sim(T_i, S^{'}) = \sum_{k \in S'} \frac{Overlap(T_i, T_k)}{Length(T_i)} \tag{3}$$

where, $Length(T_i)$ is the number of keywords of T_i. We follow Khan et al. [20] to identify the keywords of T_i as the nouns, verbs, adjectives present in T_i and similarly, for $S^{'}$, we consider the distinct set of nouns, verbs, adjectives present in all the tweets of $S^{'}$. Therefore, a lower $Sim(T_i, S^{'})$ ensures T_i has minimum redundant content information with respect to already generated summary, $S^{'}$, and a higher $E(T_i)$ ensures T_i has higher information coverage of the category.

3. Experiments and Results

In this Section, we provide details of the experiment and results. For the datasets, we consider *Los Angeles International Airport Shooting* [1] (D_1) provided by Olteanu et al. [21], *Hurricane Matthew* [2] (D_2), *Puebla Mexico Earthquake* [3] (D_3), *Pakistan Earthquake* [4] (D_4) and *Midwestern U.S. Floods* [5] (D_5) provided by Alam et al. [22] and *Sandy Hook Elementary School Shooting* [6] (D_6) provided by Dutta et al. [7]. We perform lemmatization, convert to lower case and remove of Twitter specific keywords [23] and retweets as pre-processing. We consider the ground truth summary provided by Garg et al. [9] for D_1-D_5 and by Dutta et al. [7] for D_6. We compare *EnDSUM* with content based [24] (B_1), graph based [7] (B_2), sub-event based [25] (B_3) and ontology based [9] (B_4) disaster summarization approaches.

Results and Discussion : We evaluate the performance of *EnDSUM* and the existing research with the ground truth summary using ROUGE-N [26] F1-score score when N=1, 2, and L. Our observations from Table 1 indicate that *EnDSUM* ensures better ROUGE-N F1-score over all baselines for D_2-D_6. The improvement is highest over B_1 baseline and lowest over B_4 baseline. *EnDSUM* performs worse than B_4 for Rouge-N scores and worse than B_1 for Rouge-2 and Rouge-L scores on D_1. Therefore, although *EnDSUM* has highly effective performance in most scenarios, it sometimes fails to resolve the vocabulary overlap across different categories in a disaster, as seen for D_1. Therefore, to resolve this, we are working towards making *EnDSUM* resilient irrespective of the vocabulary diversity by considering semantic and contextual similarity along with the already considered content similarity for entropy and diversity calculation.

Table 1
F1-score of ROUGE-1, ROUGE-2 and ROUGE-L score of *EnDSUM* and baselines on 6 datasets is shown.

Dataset	Approaches	ROUGE-1	ROUGE-2	ROUGE-L	Dataset	Approaches	ROUGE-1	ROUGE-2	ROUGE-L
		F1-score	F1-score	F1-score			F1-score	F1-score	F1-score
D_1	$EnDSUM$	0.55	0.21	0.27	D_4	$EnDSUM$	**0.51**	**0.16**	**0.24**
	B_1	0.49	0.22	0.29		B_1	0.20	0.04	0.20
	B_2	0.48	0.18	0.25		B_2	0.47	0.14	0.21
	B_3	0.52	0.21	0.23		B_3	0.45	0.11	0.21
	B_4	**0.56**	**0.23**	**0.29**		B_4	0.50	0.15	0.23
D_2	$EnDSUM$	**0.52**	**0.17**	**0.24**	D_5	$EnDSUM$	**0.52**	**0.13**	**0.24**
	B_1	0.48	0.13	0.22		B_1	0.19	0.04	0.18
	B_2	0.47	0.14	0.22		B_2	0.48	0.10	0.20
	B_3	0.44	0.12	0.22		B_3	0.50	0.12	0.22
	B_4	0.49	0.15	0.23		B_4	0.51	0.13	0.22
D_3	$EnDSUM$	**0.52**	0.14	**0.26**	D_6	$EnDSUM$	**0.55**	**0.27**	**0.44**
	B_1	0.45	0.13	0.23		B_1	0.53	0.26	0.33
	B_2	0.46	0.14	0.24		B_2	0.52	0.22	0.29
	B_3	0.44	0.14	0.23		B_3	0.48	0.20	0.27
	B_4	0.48	**0.16**	0.25		B_4	0.51	0.20	0.29

[1] https://en.wikipedia.org/wiki/2013_Los_Angeles_International_Airport_shooting
[2] https://en.wikipedia.org/wiki/Hurricane_Matthew
[3] https://en.wikipedia.org/wiki/2017_Puebla_earthquake
[4] https://en.wikipedia.org/wiki/2019_Kashmir_earthquake
[5] https://en.wikipedia.org/wiki/2019_Midwestern_U.S._floods
[6] https://en.wikipedia.org/wiki/Sandy_Hook_Elementary_School_shooting

4. Conclusions and Future Works

In this paper, we propose a novel entropy and diversity based tweet summarizer, *EnDSUM* for disaster events. Our experimental analysis on 6 disaster datasets indicates both the effectiveness of *EnDSUM* and its scope of improvement. For example, to handle the the high vocabulary overlap among categories, we are working to both include semantic and contextual similarity while calculating entropy and diversity in *EnDSUM*. Furthermore, while most summarization algorithms generate a predefined length summary, we intend to extend *EnDSUM* such that it provides complete information coverage of the disaster event, while maintaining diversity automatically without predefined summary length. Currently, *EnDSUM* selects the most informative tweet into the summary at each iteration. As a future direction, we intend to modify *EnDSUM* such that it can select the best subset of tweets simultaneously as the summary.

References

[1] M. Imran, P. Mitra, C. Castillo, Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages, arXiv preprint arXiv:1605.05894 (2016).

[2] C. Castillo, Big crisis data: social media in disasters and time-critical situations, Cambridge University Press, 2016.

[3] R. Chakraborty, A. Kharat, A. Khatua, S. K. Dandapat, J. Chandra, Predicting tomorrow's headline using twitter deliberations., in: CIKM Workshops, 2018.

[4] S. Vieweg, C. Castillo, M. Imran, Integrating social media communications into the rapid assessment of sudden onset disasters, in: International Conference on Social Informatics, Springer, 2014, pp. 444–461.

[5] M. Imran, C. Castillo, F. Diaz, S. Vieweg, Processing social media messages in mass emergency: A survey, ACM Computing Surveys (CSUR) 47 (2015) 1–38.

[6] K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, S. Ghosh, 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, 2015, pp. 583–592.

[7] S. Dutta, V. Chandra, K. Mehra, A. K. Das, T. Chakraborty, S. Ghosh, Ensemble algorithms for microblog summarization, IEEE Intelligent Systems 33 (2018) 4–14.

[8] A. Dusart, K. Pinel-Sauvagnat, G. Hubert, Tssubert: Tweet stream summarization using bert, arXiv preprint arXiv:2106.08770 (2021).

[9] P. K. Garg, R. Chakraborty, S. K. Dandapat, Ontorealsumm: Ontology based real-time tweet summarization, arXiv preprint arXiv:2201.06545 (2022).

[10] K. Rudra, N. Ganguly, P. Goyal, S. Ghosh, Extracting and summarizing situational information from the twitter social media during disasters, ACM Transactions on the Web (TWEB) 12 (2018) 1–35.

[11] S. Dutta, S. Ghatak, M. Roy, S. Ghosh, A. K. Das, A graph based clustering technique for tweet summarization, in: 2015 4th international conference on reliability, infocom technologies and optimization (ICRITO)(trends and future directions), IEEE, 2015, pp. 1–6.

[12] A. Khurana, V. Bhatnagar, Investigating entropy for extractive document summarization, Expert Systems with Applications 187 (2022) 115820.
[13] G. Feigenblat, H. Roitman, O. Boni, D. Konopnicki, Unsupervised query-focused multi-document summarization using the cross entropy method, in: Proceedings of the 40th International ACM SIGIR Conference on research and development in information retrieval, 2017, pp. 961–964.
[14] S. Aji, R. Kaimal, Document summarization using positive pointwise mutual information, AIRCC's International Journal of Computer Science and Information Technology 4 (2012) 47–55.
[15] W. Luo, F. Zhuang, Q. He, Z. Shi, Effectively leveraging entropy and relevance for summarization, in: Asia Information Retrieval Symposium, Springer, 2010, pp. 241–250.
[16] G. Ravindra, N. Balakrishnan, K. Ramakrishnan, Multi-document automatic text summarization using entropy estimates, in: International Conference on Current Trends in Theory and Practice of Computer Science, Springer, 2004, pp. 289–300.
[17] R. Chakraborty, M. Bhavsar, S. K. Dandapat, J. Chandra, Tweet summarization of news articles: An objective ordering-based perspective, IEEE Transactions on Computational Social Systems 6 (2019) 761–777.
[18] R. Chakraborty, M. Bhavsar, S. Dandapat, J. Chandra, A network based stratification approach for summarizing relevant comment tweets of news articles, in: International Conference on Web Information Systems Engineering, Springer, 2017, pp. 33–48.
[19] C. Hark, A. Karcı, Karcı summarization: A simple and effective approach for automatic text summarization using karcı entropy, Information Processing & Management 57 (2020) 102187.
[20] M. A. H. Khan, D. Bollegala, G. Liu, K. Sezaki, Multi-tweet summarization of real-time events, in: 2013 International Conference on Social Computing, IEEE, 2013, pp. 128–133.
[21] A. Olteanu, S. Vieweg, C. Castillo, What to expect when the unexpected happens: Social media communications across crises, in: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, 2015, pp. 994–1009.
[22] F. Alam, U. Qazi, M. Imran, F. Ofli, Humaid: Human-annotated disaster incidents data from twitter with deep learning benchmarks, arXiv preprint arXiv:2104.03090 (2021).
[23] C. Arachie, M. Gaur, S. Anzaroot, W. Groves, K. Zhang, A. Jaimes, Unsupervised detection of sub-events in large scale disasters, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 354–361.
[24] K. Rudra, P. Goyal, N. Ganguly, M. Imran, P. Mitra, Summarizing situational tweets in crisis scenarios: An extractive-abstractive approach, IEEE Transactions on Computational Social Systems 6 (2019) 981–993.
[25] K. Rudra, P. Goyal, N. Ganguly, P. Mitra, M. Imran, Identifying sub-events and summarizing disaster-related information from microblogs, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 265–274.
[26] C.-Y. Lin, Rouge: A package for automatic evaluation of summaries, in: Text summarization branches out, 2004, pp. 74–81.