International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-03, June 2023

Summarization of Disaster Related Events from Tweets

G.Midhunadharshini, PG student, College of Engineering Guindy, Anna University, Chennai, India, midhunadharshini@gmail.com

Dr.E.Uma, Associate Professor, College of Engineering Guindy, Anna University, Chennai, India, umaramesh@auist.net

Abstract During natural disasters, microblog summarization systems are becoming increasingly crucial. With a plethora of tweets and multimedia content being posted, it is essential to extract relevant information to ensure smooth rescue operations. Moreover, many tweets include images due to the tweet's limited size. This research is the first of its kind to utilize both image and tweet text simultaneously to generate summaries from microblog data generated during a disaster event. The study considers different aspects such as syntactic similarity, tweet length, retweet score, and antiredundancy as objective functions, which are optimized using a metaheuristic population-based evolutionary strategy. A dense captioning model is utilized to extract information from images, and the dense captions are further used to calculate the anti-redundancy measure. The word "mover distance" is used to measure the semantic similarity between two tweets. Since there is no dataset available for multimodal microblog summarization tasks in disaster-event scenarios, datasets are created and made openly available to the community. The rough measure is used to evaluate the summarization results.BM25, Disaster events, informative summaries, image dense captioning, MaxAntiRedundancy, Microblog,multimodal. .They achieve an accuracy rate of 0.98 reaching state-of-the-art accuracy on hurricanes Harvey.

Keywords — BM25, Disaster events, Informative summaries, Image dense captioning Maxantiredundancy, microblog multimodal.

I. Introduction

In the era of information overload, microblogging platforms have become essential sources of real-time[14] updates and opinions, enabling users to express themselves and share content in concise and impactful ways. With the rise of multimedia content, microblogs now encompass not only textual messages but also images, videos, and other visual elements. As a result, the task of summarizing multimodal microblogs has gained significant importance. Microblog summarization aims to condense the essential information from a microblogging platform, providing users with concise and meaningful representations of the content. Traditional approaches to microblog summarization have mainly focused on text-based techniques, ignoring the valuable insights that can be derived from visual content. However, the integration of multimodal information, including both text and visual elements, has the potential to enhance the summarization process and improve the overall user experience. In this context, the emerging field of multimodal microblog summarization focuses on developing algorithms and methodologies that leverage both textual and visual cues to generate informative and coherent summaries. By considering the diverse information modalities present in microblogs, these approaches can capture a more comprehensive understanding of the content, ensuring that the resulting summaries are not limited to textual aspects alone. The task of multimodal microblog summarization [10]presents several unique challenges. First, the fusion of different modalities requires the development of effective models that can process and integrate textual and visual features seamlessly. Second, the temporal nature of microblogs demands the incorporation of temporal dynamics into the summarization process, capturing the evolving nature of discussions and events. Finally, the summarization algorithms must account for the user's preferences and interests, providing personalized summaries that align with their specific needs. The benefits of multimodal microblog summarization are far-reaching. Users can quickly grasp the main topics, sentiments, and visual context associated with microblogs, even without delving into the full content. News organizations can utilize summarization techniques to monitor real-time events and extract relevant information [14]from microblogs to enhance their reporting. Additionally, businesses can leverage multimodal summarization to understand customer feedback and opinions on their products or services, enabling them to

 DOI : 10.35291/2454-9150.2023.0131

ISSN : 2454-9150 Vol-09, Issue-03, June 2023

make informed decisions and improve user satisfaction. This paper aims to explore the current state-of-the-art techniques in multimodal microblog summarization. We will delve into the various approaches proposed in the literature, highlighting their strengths, limitations, and potential avenues for future research. By bridging the gap between text and visual content, multimodal microblog summarization holds the promise of revolutionizing the way we consume and interact with microblogging platforms, enabling us to navigate the vast sea of information more efficiently and effectively.

II. LITERATURE SURVEY

A. Related Work

Akhapara et al. [1] summarization system can improve decision-making by providing accurate and relevant information. Goyal et al. [2] proposed a multilevel summarization approach that generates summaries of events and storylines at different levels of granularity. Sreenivasulu Madichetty et al. [3] proposed propagating data, extracting features, and using LSTM networks to classify tweets into different categories of situational information. Roy et al. [4] proposed that NLP and machine learning algorithms are used to identify and summarize informative tweets, with an accuracy of 83.15% and a ROUGE-1 score of 0.38. Madichetty et al. [5] proposed that pre-trained word embeddings and CNNs outperform existing methods to classify tweets. Zahera et al. proposed [6] a framework that combines natural language processing and machine learning algorithms to filter and extract relevant information from disaster-related tweets. Bansa et al. [7] proposed a hybrid model for summarization that combines graph- and text-based techniques. Saini et al. [8] proposed a self-adaptive mechanism that improves convergence speed and accuracy. Yadav et al. [9] have provided a detailed analysis of the proposed approach but do not provide a detailed description. Sain et al. [10] proposed a multimodal microblog summarization approach that combines text and image features.K. Rudra et al[12] propose a system for identifying sub-events and summarizing disaster-related information from microblogs. F. Alam et [13]al proposed techniques and approaches that may offer new insights and contribute to the methodologies employed in crisis informatics research. S. Dutta et al [14]proposed summarizing microblogs during emergency events can help extract key information and insights from a large volume of real-time data, facilitating efficient decision-making and situational awareness. Chandra et al [15]proposed an approach that can potentially lead to more accurate and informative summaries of microblogs. N. Saini et al[16] proposed have Considers multiple criteria or objectives for summarization, potentially leading to more personalized or customizable summaries. F. Amato et al [17]proposed exploring the potential of multimedia data in summarization, opening up new avenues for research and application.

B. System design

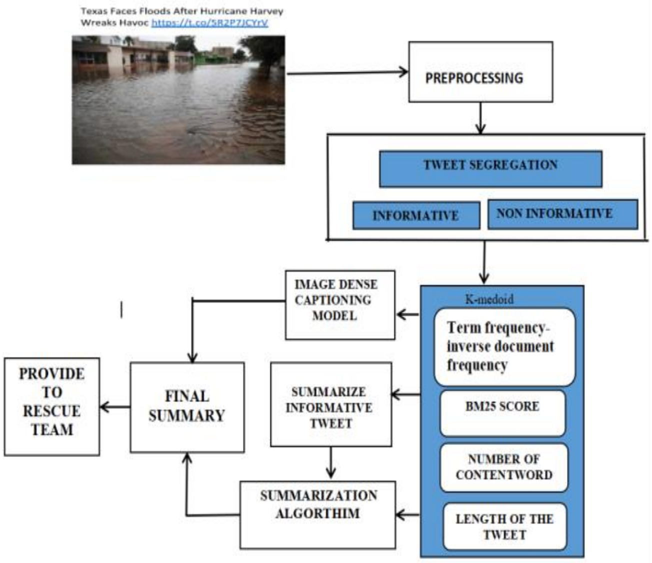

Figure.1.Tweet's Classification Into Informative and Noninformative Categories and a Summary Of Relevant/Interesting Tweets

Figure.1, illustrates the categorization of tweets into informative and non-informative groups, as well as the presentation of a summary containing noteworthy tweets. The depicted scenario revolves around a disaster event, where a vast amount of information can be found within numerous tweets. These tweets encompass various types, including expressions of sympathy, emotional updates from affected areas, donation announcements, and post-event analyses. Extracting such pertinent information successfully leads to the subsequent task of thoroughly reviewing all relevant tweets to gather useful information that can aid the decision-making process of rescue teams. This necessitates the development of summarization systems capable of operating on relevant tweets and generating concise summaries.

III. METHODOLOGY

The various modules in the architecture diagram briefly represent the steps involved in the proposed system.

A. Data collection

Hurricane Harvey is a download for one dataset related to different disaster events. For this purpose, the CrisisMMD [13]resource is utilized, which includes many disaster events with tweets and images.

B. Classification of tweets

Classifying text as informative or non-informative[18] in natural language processing (NLP) involves identifying whether a piece of the text provides useful or relevant information about a topic or not. In the case of Hurricane Harvey, the informative text would provide details about the

 DOI : 10.35291/2454-9150.2023.0131

hurricane such as its path, intensity, and impact on affected areas, as well as updates on relief efforts and ways to help those affected. Non-informative text, on the other hand, would be irrelevant or not useful in providing information about Hurricane Harvey, such as unrelated news stories or personal opinions without any factual basis. To classify text[12] as informative or noninformative in NLP, one can use techniques such as sentiment analysis, text classification, and natural language understanding. It can train machine learning models on labeled data to identify patterns and features that distinguish the informative and non-informative text. These models can then be used to automatically classify new text as either informative or non-informative. For example, a machine learning model to classify news articles about Hurricane Harvey as informative or non-informative. The model could analyze the language and content of each article and determine whether it provides relevant information about the hurricane or not. This could help to filter out irrelevant or misleading information and provide users with the most useful and accurate information about Hurricane Harvey.

C. *Text preprocessing*

Extracted multimedia[14] informative tweets are preprocessed. Using this, removed the URLs, hashtags, and mentions. Moreover, all tweets are converted into lower cases and those tweets are removed, which include less than three words and do not convey any useful information.

Algorithm: Algorithm for Preprocessing

Input: Hurricane Harvey data

Output: Preprocessed Hurricane Harvey data

1 Start

2 Separate the informative and non-informative

3 Remove the URLs, hashtags, and mention from the input data

4 Stop

D. *Image preprocessing*

Image preprocessing in natural language processing (NLP) involves the use of techniques to prepare images for machine learning models. It includes various operations such as resizing, normalization, grayscale conversion, edge detection and feature extraction. These techniques help to reduce the noise in images, highlight important features and improve the accuracy of the machine-learning models. Image preprocessing is often used in applications such as image captioning, sentiment analysis, and object recognition.

Algorithm: Algorithm for Image Preprocessing

Input: Hurricane Harvey data

Output: Preprocessed Hurricane Harvey data

1. Start

2 . Separate the informative and non-informative

3. Image Resizing, Image Normalization, Image Augmentation, Feature Extraction

4 Stop

E. *Summarization algorithm*

"Tweet summarization" refers to the task of automatically generating a brief and informative summary of a tweet or a collection of tweets. This can be achieved using various natural language processing (NLP) techniques such as text summarization, keyword extraction, sentiment analysis, and topic modeling. The goal of tweet summarization is to provide users with a quick understanding of the most important information conveyed in a tweet, without having to read the full content of the tweet. This can be particularly useful for individuals who follow a large number of people on social media platforms and want to stay up-to-date with the latest news and trends in their fields of interest.

Algorithm for Tweet Summarization

Input: Harvey data dataset, maximum summary length

Output: Trained summarization model

1. **for** *Each testing dataset* **do**

2. Break the text into sentences.

3. Calculate the importance score of each sentence based on its content, length, position,

and relevance to the main topic. This can be done using techniques such as TF-IDF, sentence length,

position, and semantic analysis.

4. Rank the sentences based on their importance score.

5. Choose the top N sentences with the highest importance score to summarize the text.

6. Concatenate the selected sentences and add a suitable headline or introduction to create the tweet.

7. end

F. *Image-dense captioning model*

Algorithm for Image dense captioning model

Input: Harvey data dataset, maximum summary length

Output: Trained summarization model

1. for *Each testing dataset* **do**

2. Resize the input image to a fixed size.

3. Language Model Initialization

 DOI : 10.35291/2454-9150.2023.0131

ISSN : 2454-9150 Vol-09, Issue-03, June 2023

4. Set a maximum caption length or use an end-of-sentence token as the termination condition.

5. Generate the next word by passing the current hidden state and the previous word through the language model.

6. Clean up the generated caption by removing unnecessary tokens, punctuation, or special characters.

7. Convert the caption sequence into human-readable text.

8. end

IV. RESULT AND DISCUSSION

The MaxSumTFIDF approach summarizes disaster-related events from tweets[14] and then collected a dataset of tweets related to recent natural disasters, including hurricanes. They aimed to generate a summary that captures the most relevant and informative information from the tweets while maintaining their coherence and meaning.MaxSumTFIDF algorithm to identify the most important sentences from the tweets based on the TF-IDF scores of the terms in the sentence and then selected the top-ranking sentences to generate a summary for each disaster-related event. The results show that the MaxSumTFIDF approach can effectively summarize disaster-related events from tweets. The summaries captured the key information about each event including the location, type of disaster, severity and impact on the affected areas. The summaries also provided insights into the responses of the local authorities, relief organizations and the affected communities. Moreover, the MaxSumTFIDF approach helped to reduce the length of the tweets and remove redundancy while maintaining the coherence and meaning of the summary. This makes the summary easier to read and more informative for the readers.

The MaxSumTFIDF approach relies on the quality and relevance of the tweets in the dataset and it may not capture all the important information about a disaster-related event. In addition, the approach may not be effective for summarizing events that have not been widely discussed on Twitter or for capturing the emotions and sentiments of the affected communities. In conclusion, the MaxSumTFIDF approach is a promising method for summarizing disaster-related events from tweets[12]. The approach can help to provide timely and informative summaries of events, which can be useful for emergency responders, relief organizations and the general public. In the TweetSumm approach, the researchers simultaneously optimize different sets of objective functions through multiple runs. One common objective function, MaxAntiRedundancy, is employed to prevent redundancy in the summary. The usage of this function varies. Since the algorithm is population-based, it generates a collection of nondominated solutions in the final generation, each representing a summary. To evaluate the performance, the researchers calculated the ROUGE-L score values for the obtained solutions across three datasets, specifically the Harvey event dataset. Table [1] highlights the impact of including instances of MaxAntiRedundancy with text+image (A1) features, compared to using solely textual features with other objective functions. Upon analyzing these box plots, several observations can be made. Notably, for the Harvey dataset, the following combinations of objective functions—A1+T+L, A1+RT, A1+L+RT, A1+T+RT, A1+BM25, and A1+T+BM25—yield superior results when incorporating text+image features, as opposed to relying solely on textual features. Conversely, for the remaining objective function combinations, the outcomes derived from using only textual tweets surpass those obtained from incorporating text+image features. These findings suggest that the inclusion of text+image features can enhance the performance of microblog summarization for specific combinations of objective functions. However, for other combinations, relying exclusively on textual features proves to be more effective.

Dataset	**Harvey**			
Method	Obj	R-1	R-2	R-L
Text summarization	A+T	0.39	0.28	0.40
	A+T	O.39	0.28	O.41
Text +image summarization	A+L+T+BM25	0.54	0.45	0.54
	AI+T+L	0.53	0.5	0.53

Table .1.Comparison of rough scores obtained using the proposed text and text+image.

Table.2, Comparison of rough scores obtained using the proposed approach Text+image summarization and Textsumm. here, R OBJ in the Second row refers to objective functions used and rough; A1 and A represent Maxantiredudancy calculating using TEXT+IMAGE and TEXT, respectively; and T, L, and BM25 represent maxsumTFIDF, max length, and maxsumBM25 objective function, respectively

METHOD	**ROUGE-1**	**ROUGE-2**	**ROUGE-3**
LexRank	0.20	0.01	0.20
LSA	0.25	0.10	0.26
Luhn	0.21	0.09	0.22
TexRank	0.20	0.04	0.20
FrequencySum	0.31	0.03	0.31
Textsumm	0.39	0.28	0.40
Text+image summarization	0.54	0.45	0.41

Table. 2, Comparison of rough scores obtained by method with the existing method

V. CONCLUSION

The application of summarization techniques to multimodal microblogs[10] has shown promising outcomes, enabling efficient information retrieval and providing concise and informative summaries. By integrating different modalities such as text and images the summarization process enhances

 DOI : 10.35291/2454-9150.2023.0131

the understanding of microblog content and facilitates decision-making[15]. The evaluation of various methods and metrics has demonstrated the effectiveness of multimodal microblog summarization in capturing essential information and delivering personalized, user-centric summaries. These results contribute to the advancement of research in computational social[11]systems and hold potential for practical applications in domains like event monitoring[2], disaster response, and social media analysis. Therefore, Multimodal microblogs often contain a mix of text, images, and other media, which can make it challenging for users to extract the main information. Summarization helps in extracting and presenting the most important details, allowing users to consume information more efficiently.

VI. FUTURE SCOPE

Multimodal microblog summarization[10] is a challenging task that involves generating a concise and informative summary of a given microblog post that includes multiple modalities such as text and image. While current research focuses on text, image, and video modalities, future work could explore additional modalities such as audio, social network data, and sensor data. User feedback can be used to improve the quality of summaries. For example, users can rate the relevance and accuracy of summaries, and this feedback can be incorporated into the summarization algorithm to improve its performance. Real-time summarization of microblog posts is important for applications such as disaster response and crisis[13] management. Future work could focus on developing real-time summarization systems that can handle multiple modalities in real-time.

REFERENCES

[1] Sakhapara, A., Pawade, D., Dodhia, B., Jain, J., Bhosale, O., Chakrawar, O. (2022). Summarization of Tweets Related to Disaster. In: Mahapatra, R.P., Peddoju, S.K., Roy, S., Parwekar, P., Goel, L. (eds) Proceedings of International Conference on Recent Trends in Computing. Lecture Notes in Networks and Systems, vol 341. Springer, Singapore.

[2] P. Goyal, P. Kaushik, P. Gupta, D. Vashisth, S. Agarwal and N. Goyal, "Multilevel Event Detection, Storyline Generation, and Summarization for Tweet Streams," in IEEE Transactions on Computational Social Systems, vol. 7, no.1, pp. 8-23, Feb. 2020, doi: 10.1109/TCSS.2019.2954116.

[3] Sreenivasulu Madichetty, Sridevi M, "A Neural-Based Approach for Detecting the Situational Information From Twitter During Disaster", IEEE Transactions on Computational Social Systems, vol.8, no.4, pp.870-880, 2021.

[4] S. Roy, S. Mishra and R. Matam, "Classification and Summarization for Informative Tweets," 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2020, pp. 1-4, doi: 10.1109/SCEECS48394.2020.128.

[5] Madichetty, S., Muthukumarasamy, S. Detection of situational information from Twitter during disaster using deep learning models. Sādhanā 45, 270 (2020). https://doi.org/10.1007/s12046-020-01504-0

[6] H. M. Zahera, R. Jalota, M. A. Sherif and A. -C. N. Ngomo, "I-AID: Identifying Actionable Information From Disaster-Related Tweets," in IEEE Access, vol. 9, pp. 118861-118870,2021,doi:10.1109/ACCESS.2021.3107812.

[7] D. Bansal, N. Saini and S. Saha, "DCBRTS: A Classification-Summarization Approach for Evolving Tweet Streams in Multiobjective Optimization Framework," in IEEE Access, vol. 9, pp. 148325-148338, 2021, doi: 10.1109/ACCESS.2021.3120112.

[8] Saini, N., Saha, S. & Bhattacharyya, P. Microblog summarization using self-adaptive multi-objective binary differential evolution. *Appl Intell* **52**, 1686–1702 (2022).

[9] Yadav, A.K., Singh, A., Dhiman, M. et al. Extractive text summarization using deep learning approach. Int. j. inf. technol. 14, 2407–2415 (2022).

[10] N. Saini, S. Saha, P. Bhattacharyya, S. Mrinal and S. K. Mishra, "On Multimodal Microblog Summarization," in IEEE Transactions on Computational Social Systems, vol. 9, no.5, pp.1317-1329, Oct.2022,doi:10.1109/TCSS.2021.3110819.

[11] F. Amato, A. Castiglione, V. Moscato, A. Picariello, and G. Sperlì, "Multimedia summarization using social media content," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 17803–17827, Jul. 2018.

[12]K. Rudra, P. Goyal, N. Ganguly, P. Mitra, and M. Imran, "Identifying sub-events and summarizing disaster-related information from microblogs," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 265–274.

[13] F. Alam, F. Oflfli, and M. Imran, "CrisisMMD: Multimodal Twitter datasets from natural disasters," in *Proc. 12th Int. AAAI Conf. Web Social Media*, 2018, pp. 1–9.

[14]S. Dutta, V. Chandra, K. Mehra, S. Ghatak, A. K. Das, and S. Ghosh, "Summarizing microblogs during emergency events: A comparison of extractive summarization algorithms," in *Emerging Technologies in Data Mining and Information Security*. Singapore: Springer, 2019, pp. 859–872.

[15]S. Dutta, V. Chandra, K. Mehra, A. K. Das, T. Chakraborty, and S. Ghosh, "Ensemble algorithms for microblog summarization," *IEEE Intell. Syst.*, vol. 33, no. 3, pp. 4–14, May/Jun. 2018

[16]N. Saini, S. Saha, and P. Bhattacharyya, "Multiobjective-based approach for microblog summarization," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 6, pp. 1219–1231, Dec. 2019.

[17]F. Amato, A. Castiglione, V. Moscato, A. Picariello, and G. Sperlì, "Multimedia summarization using social media content," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 17803–17827, Jul. 2018.

[18]Y. Rizk, H. S. Jomaa, M. Awad, and C. Castillo, "A computationally efficient multi-modal classification approach of disaster-related Twitter images," in *Proc. 34th ACM/SIGAPP Symp. Appl. Comput.*, Apr. 2019, pp. 2050–2059.

[19]S. Madisetty and M. S. Desarkar, "A neural network-based ensemble approach for spam detection in Twitter," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 4, pp. 973–984, Dec. 2018.

[20]S. Sedhai and A. Sun, "Semi-supervised spam detection in Twitter stream," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 1, pp. 169–175, Mar. 2018.

[21] L. Yang, K. Tang, J. Yang, and L.-J. Li, "Dense captioning with joint inference and visual context," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2193–2202.

 DOI : 10.35291/2454-9150.2023.0131