Generalized News Event Discovery via Dynamic Augmentation and Entropy Optimization (Supplementary Materials)

Anonymous Authors

A SUPPLEMENTARY DETAILS OF THE MULTIMODAL NEWS EVENT DISCOVERY (MNED) DATASET

A.1 Detailed Statistics of the MNED Dataset

As shown in Figure 1 and Figure 2, detailed data statistics are provided. We observe that the number of events for long-term and cyclical events is significantly higher compared to short-term events, because they contain many sub-events. By categorizing and collecting these three types of events, our dataset is not only more closely with real-world scenarios but also better suited for the task of generalized news event discovery.

A.2 Comparisons with Existing Datasets

Table 1 compares our MNED dataset with other existing datasets. From this comparison, we have the following observations:

- Most current datasets do not include temporal metadata, which is a crucial attribute for news event discovery tasks. This omission is typically because these datasets are designed for the closed-setting event discovery, where training and test sets are randomly split based on event categories rather than by time. In our work, we advocate for splitting training and test sets based on chronological order, which more accurately reflects the temporal nature of real-world news event discovery tasks.
- Most existing datasets have relatively few samples, which is not conducive to learning news event features effectively. Events, representing complex semantic entities, require a substantial number of samples to capture their nuances fully. For the SED dataset, it includes a large volume of data but is limited to coarse-grained event labels such as parties and festivals. In contrast, our dataset not only provides a large number of more fine-grained event categories but also categorizes these events into short-term, cyclical, and long-term events. This categorization is beneficial for models to learn distinctive features associated with different types of events.
- The majority of existing datasets predominantly consist of English posts. This is because the collection process intentionally filters out other languages to simplify the analysis. However, news event discovery tasks inherently involve events from diverse countries, implying that multiple languages are common and that the local language of the event can offer a more authentic perspective for interpreting the event. Thus, our dataset retains posts in various languages, which, while increasing the complexity of the task, also provides multiple viewpoints that aid the model in understanding the event more comprehensively.
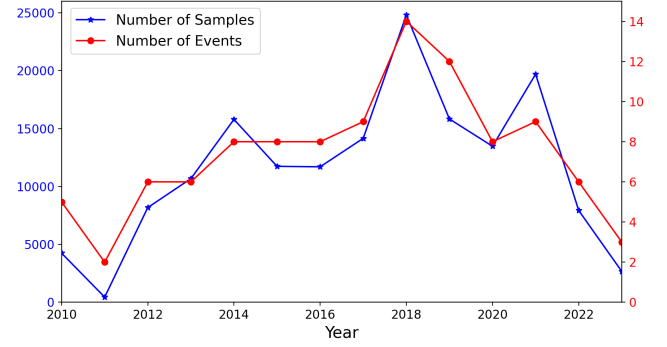

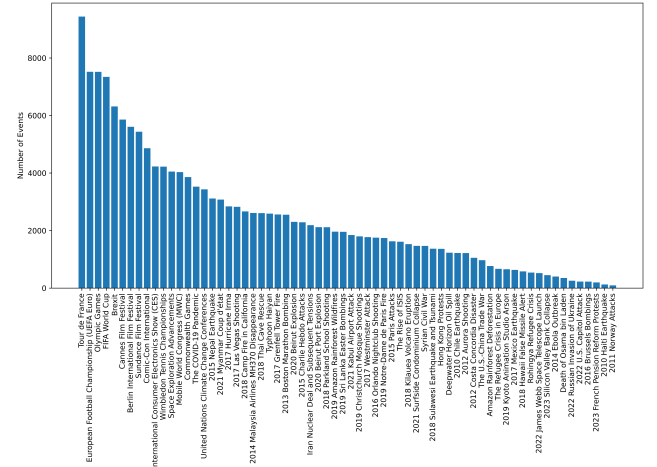Figure 1: Statistics on the number of samples and events in the MNED dataset.

Figure 2: Statistics for all events in the MNED dataset.

B SUPPLEMENTARY EXPERIMENTS ON THE EXISTING DATASET

To validate the generalization capability of our proposed Dynamic Augmentation and Entropy Optimization (DAEO) model, we conduct experiments on a public dataset, i.e., the CrisisMMD dataset [1].

B.1 CrisisMMD Dataset

This dataset is a multimodal crisis dataset that encompasses seven natural disaster events from 2017, including Hurricane Irma, Hurricane Maria, Hurricane Harvey, the Mexico earthquake, the Iraq–Iran earthquakes, the Sri Lanka floods, and the California wildfires. Detailed statistics of the dataset are shown in Table 1.

Table 1: Comparison of existing datasets. ("#" represents the number of samples.)

Dataset	Platform	#Sample	#Event	Modality	Fine-grained	Temporal Metadata	Multilanguage	Public
CE [13]	Twitter	800	2	Single	no	yes	no	no
SED [15]	Flickr, Youtube	427,370/1,327	21,169	Multiple	no	yes	no	yes
ASO [17]	Twitter	1,100	3	Single	no	no	no	no
OSMNs [3]	Twitter	3.5M	20	Single	no	yes	no	no
Twevent [7]	Wikipedia,Twitter	3.2M/4.3M	N.A.	Single	no	no	no	no
DHS [6]	Twitter, Tumblr	2.1M/0.3M	600	Multiple	no	no	no	no
PHEME [22]	Twitter	2,089	9	Multiple	yes	no	no	yes
NED [9]	Twitter	17,366	40	Multiple	yes	no	no	yes
CrisisMMD [1]	Twitter	18,126	7	Multiple	yes	no	no	yes
MNED	Twitter	161,350	66	Multiple	yes	yes	yes	yes

Table 2: Experimental results on the CrisisMMD dataset with closed setting (without new events).

Measure	f-CLSWGAN [20]	TCGAN [12]	CADA-VAE [16]	DAVAE [5]	MDL-DR [11]	Multi-RC [21]	SCBD [1]	AT-CVAE [8]	OWSEC [14]	DAEO
Accuracy	0.7582	0.8954	0.7412	0.7977	0.8677	0.8395	0.9366	0.9718	0.9672	**0.9722**
Macro F1	0.7578	0.8936	0.7406	0.7873	0.8573	0.8223	0.9510	0.9709	0.9709	**0.9758**

Table 3: The division of the CrisisMMD dataset. '#New' refers to the number of new events.

Training set		Test set		
#Sample	#Event	#Sample	#Event	#New
6,047	3	10,567	7	4

B.2 Data Partitioning

Given the analysis in A.2, the CrisisMMD dataset lacks temporal information, which is used for closed-setting event discovery [8]. Therefore, we are not able to divide the training and test sets according to time for the generalized news event discovery task. Following [18], we extract the first three categories of events as known events from the dataset. We set 50% of the data from these categories for the training set. The remaining 50% of data from these categories, along with all event samples from the other four categories, constitute the test set. For the validation set, we select 20% of the training set, chosen randomly across categories. The specific partitioning details are depicted in Table 3.

B.3 Performance on the CrisisMMD Dataset

Table 4 shows the experimental results of our DAEO model on the CrisisMMD dataset. From these results, we observe the following:

- Our model achieves the best performance on this public dataset compared to other methods, which validates its strong generalization ability.
- Our model exhibits high accuracy on known classes on the CrisisMMD dataset. This is partly due to the use of random partitioning to define known and unknown events, which simplifies the task to some extent. This result also underscores the importance of partitioning training and test sets based on time to prevent potential future information leakage, which is crucial for realistic event discovery tasks.

Table 4: Results on the CrisisMMD dataset.

Method	Known	New	All
K-means [10]	0.315	0.526	0.375
RankStats [4]	0.854	0.425	0.732
UNO [2]	0.946	0.531	0.828
GCD [18]	0.391	0.462	0.363
SimGCD [19]	0.962	0.610	0.862
DAEO	**0.968**	**0.669**	**0.883**
Δ	+0.006	+0.060	+0.021

- Our model also performs well on new categories, indicating that the features generated by the proposed multimodal augmentation module are robust even for new events.

B.4 Performance on the CrisisMMD Dataset under Closed Setting

To validate the effectiveness of our proposed multimodal augmentation module in generating robust features, we also compare its performance in a closed-setting event discovery scenario. Following [8], we divide 70% of the CrisisMMD dataset as the training set, 10% as the validation set, and 20% as the test set. The evaluation metrics used are accuracy and macro-averaged F1 score.

As shown in Table 2, we have the following observations:

- Our DAEO model outperforms other event discovery methods, which can be attributed to our multimodal augmentation module that employs adversarial techniques. The adversarial approach in feature generation effectively enhances the variability and representational capacity of the features, which in turn improves the classifier's ability to discriminate between different event types accurately.
- Combined with Table 4, we observe that the accuracy for known events in our model under the generalized setting remains very close to the accuracy under a closed setting, even after the addition of new events. This is attributed to our

adaptive entropy optimization strategy, which selectively optimizes for both known and new events. By maintaining accuracy for known events while encouraging exploration of new events, this strategy ensures that the model remains effective across all categories without compromising its ability to identify events it has previously learned.

REFERENCES

[1] Firoj Alam, Ferda Ofli, and Muhammad Imran. 2018. Crisismmd: Multimodal twitter datasets from natural disasters. In *Proceedings of the international AAAI conference on web and social media*, Vol. 12.

[2] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. 2021. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9284–9292.

[3] Hansu Gu, Xing Xie, Qin Lv, Yaoping Ruan, and Li Shang. 2011. Etree: Effective and efficient event modeling for real-time online social media networks. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 300–307.

[4] Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. 2021. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 10 (2021), 6767–6781.

[5] Mengmeng Jing, Jingjing Li, Lei Zhu, Ke Lu, Yang Yang, and Zi Huang. 2020. Incomplete cross-modal retrieval with dual-aligned variational autoencoders. In *Proceedings of the 28th ACM international conference on multimedia*. 3283–3291.

[6] Satya Katragadda, Ryan Benton, and Vijay Raghavan. 2017. Framework for real-time event detection using multiple social media sources. (2017).

[7] Chenliang Li, Aixin Sun, and Anwitaman Datta. 2012. Twevent: segment-based event detection from tweets. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 155–164.

[8] Zhangming Li, Shengsheng Qian, Jie Cao, Quan Fang, and Changsheng Xu. 2022. Adaptive transformer-based conditioned variational autoencoder for incomplete social event classification. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1698–1707.

[9] Zehang Lin, Jiayuan Xie, and Qing Li. 2024. Multi-modal news event detection with external knowledge. *Information Processing & Management* 61, 3 (2024), 103697.

[10] James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. Oakland, CA, USA, 281–297.

[11] Ferda Ofli, Firoj Alam, and Muhammad Imran. 2020. Analysis of social media data using multimodal deep learning for disaster response. *arXiv preprint arXiv:2004.11838* (2020).

[12] Frederik Pahde, Mihai Puscas, Tassilo Klein, and Moin Nabi. 2021. Multimodal prototypical networks for few-shot learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2644–2653.

[13] Ana-Maria Popescu and Marco Pennacchiotti. 2010. Detecting controversial events from twitter. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 1873–1876.

[14] Shengsheng Qian, Hong Chen, Dizhan Xue, Quan Fang, and Changsheng Xu. 2023. Open-world social event classification. In *Proceedings of the ACM Web Conference 2023*. 1562–1571.

[15] Timo Reuter, Symeon Papadopoulos, Giorgos Petkos, Vasileios Mezaris, Yiannis Kompatsiaris, Philipp Cimiano, Christopher De Vries, and Shlomo Geva. 2013. Social event detection at mediaeval 2013: Challenges, datasets, and evaluation. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop Barcelona, Spain, October 18-19, 2013*. Citeseer.

[16] Edgar Schonfeld, Sayna Ebrahimi, Samarth Sinha, Trevor Darrell, and Zeynep Akata. 2019. Generalized zero-and few-shot learning via aligned variational autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8247–8255.

[17] Samir Tartir and Ibrahim Abdul-Nabi. 2017. Semantic sentiment analysis in Arabic social media. *Journal of King Saud University-Computer and Information Sciences* 29, 2 (2017), 229–233.

[18] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. 2022. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7492–7501.

[19] Xin Wen, Bingchen Zhao, and Xiaojuan Qi. 2023. Parametric classification for generalized category discovery: A baseline study. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16590–16600.

[20] Yongqin Xian, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. 2018. Feature generating networks for zero-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5542–5551.

[21] Li Xukun and Doina Caragea. 2020. Improving disaster-related tweet classification with a multimodal approach. In *ISCRAM 2020 conference proceedings–17th international conference on information Systems for Crisis Response and Management*.

[22] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting context for rumour detection in social media. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9*. Springer, 109–123.