FUSED: Fusing Social Media Stream Classification Techniques for Effective Disaster Response

Viyom Mittal, Hongmiao Yu, K. K. Ramakrishnan
Computer Science and Engineering Department
University of California, Riverside
vmitt003@ucr.edu, hyu125@ucr.edu, kk@cs.ucr.edu

***Abstract*—Timely delivery of the right information to the right first responders can help improve the outcomes of their efforts and save lives. With social media communications (Twitter, Facebook, etc.) being increasingly used to send and get information during disasters, forwarding them to the right first responders in a timely manner can be very helpful. We use Natural Language Processing and Machine Learning, to steer the social media posts to the most appropriate first responder.**

An important goal is to retrieve and deliver only the critical, actionable information to first responders in real-time. We examine the overall pipeline starting from retrieving tweets from the social media platforms, to their classification, and dissemination to first responders.

We propose improvements in the area of data retrieval, relevance prediction and prioritizing information sent to the first responders by fusing NLP and ML classification techniques thus improving emergency response. We demonstrate the effectiveness of our proposed approach in retrieving and extracting 37,295 actionable tweets related to the IDA hurricane that occurred in the US in Aug.-Sep, 2021.

***Index Terms*—Disaster management, Data retrieval, Data analysis, Classification**

I. Introduction

Timely delivery of the right information to the right first responders as part of an emergency response effort can help improve the outcomes of their efforts and save lives. Social media communications (Twitter, Facebook, etc.) are being increasingly used to send and get information during disasters, both during natural and man-made disasters. People post critical information on these social media platforms which, if responded to in a timely manner can prove to be life saving. Forwarding the information to the right first responders in a timely manner can be very helpful. We had proposed FLARE [1], a framework for efficient and timely dissemination of relevant posts to the right first responders. In this work we develop a mechanism to efficiently fetch tweets from Twitter, a widely used social media platform. FLARE used 'Social Media Engines' (SMEs) to map social media posts (SMPs), such as tweets, to the right names. FLARE used natural language processing-based classification and then a name-based publish/subscribe framework to deliver tweets to subscribing first responders in an online real-time manner.

This paper enhances the classifier for FLARE, in an effort to build a deployable end-to-end framework to mitigate disaster and deal with its aftermath. We start with improving the data retrieval approach from the Social media platforms such as Twitter. The traditional keyword based retrieval of tweets can result in excessive amount of data and most of the this data is unrelated to the event. To avoid this case we first apply keyword based fetching of tweets only from the disaster affected regions. We leverage the text in these tweets to find the most frequently used keywords and hashtags which are specific to the particular event. These keywords are used to again fetch the tweets for collecting majority of the relevant information.

After the data retrieval, the next big challenge is finding the relevant information from the dataset. For this purpose, traditional FLARE made use of C1 classifier which predicted whether the incoming post was relevant to the incident or not. There can be different types of information which may be relevant to the incident but may not be suitable for sending to the first responders. We upgrade this C1 classifier to classify the post into one or more of the 25 different classes listed in Table I. This gives the Incident commander the flexibility to decide that out of the 24 different relevant classes, the ones that are suitable for forwarding to the first responders. Since we keep the model design similar to FLARE, it is lightweight and incurs less communication and computation overhead during the Federated learning [2] procedure.

Another major challenge in classifying the disaster related posts is the handling of class-imbalance. In the dataset, the number of critical posts that require immediate assistance is very low. We need to have a mechanism to make sure we do not miss these posts during the classification. For this purpose, we select the optimum value for the threshold used for deciding whether a post belongs to a particular class or not. We make use of Youden's J statistic [3] to calculate these threshold values.

We show the applicability of these techniques by fetching the tweets for a recent disaster, the IDA hurricane that occurred on the east coast of US in Aug.-Sep, 2021. We then use it to test our updated C1 classifier. Finally, we show the importance of selecting optimum threshold values to catch the rare critical tweets. The main contributions of the paper are as follows:

1) We develop a keyword based approach to fetch data from the Twitter social media platform without introducing excessive irrelevant data and at the same time collecting a majority of the relevant information.

2) We propose an improved classification model to find the *actionable* items from social media posts instead of just categorizing the tweet as either 'relevant' or 'irrelevant'.

3) We improve the model by updating the thresholding mechanism for getting better performance in classifying tweets into minority classes.

II. Background

During disasters, delivering rescue messages to first responders quickly and precisely is critical to improve the effectiveness of emergency response. Previous work has shown that social media has an important role to play during disasters [4] [5], because people frequently share emergency information on social media [6]. Thus, delivering emergency information posted on social media to the first responders can be very helpful.

Authorities depend on pre-defined procedures for managing disaster situations, often in the form of task assignment for personnel assigned to specific roles customized for a specific disaster. In our previous work [7], we have developed a flexible role-based dynamic naming framework in conjunction with a publish/subscribe framework for timely dissemination information to the right first responders. First responders subscribe to the appropriate names for receiving information relevant to their roles. The namespace is leveraged by FLARE [1], a framework to map social media post to right disaster management namespace. As shown in Fig. 1, FLARE proposes an end-to-end framework for fetching disaster related posts from Social media platforms, extracting the critical information and delivering them to the right first responders in real-time. FLARE is able to provide efficient and timely dissemination of relevant information to the correct first responders while utilizing the specialized knowledge of the dispatchers belonging to various participating departments. For this purpose, it sets up separate Social Media Engine for each department which are assisted by dispatcher for the learning process share their knowledge over time. Each Media Engine uses three classifiers: a C1 classifier to determine whether a social media post is relevant to the current incident; a C2 classifier that maps disaster-relevant messages to the right department or organization name. Finally, a C3 classifier maps social media posts to fine-grained prefixes. In this paper, focus on providing an improved, fine-grained C1 classifier to go beyond the simple 'relevant'/'irrelevant' binary classification.

Many works have studied fetching disaster-related Twitter posts in emergency management. Kiran et. al. [8] give a method to identify eyewitness messages during disasters. Banujan et. al. [9] fetch tweets, and delete unrelated tweets, based on keywords. Kabir et. al. [10] provide a way to fetch and process tweets based on keywords and location in real time. However, both keyword-based datasets and location-based datasets have limitation [11]. Geo-tagged tweets are very noisy and inaccurate [12]. In our work, we develop a 2-step tweet-fetching method combining geo-tags and keywords to get the best of both approaches.

Text REtrival Conference Incident Streams (TREC-IS) [13] is a TREC track is dedicated to research technologies for automatically processing social media streams during emergency situations. TREC-IS provides a large Twitter dataset collected from a range of past disasters. The dataset contains 75 events, comprising 90,000 tweets and assigned over 185,000 labels. The tweets in the dataset were manually labeled into 25 information types and 5 priority types. This detailed classification gives Incident commander the flexibility to choose among the different relevant categories he/she decide to forward to the first responders.

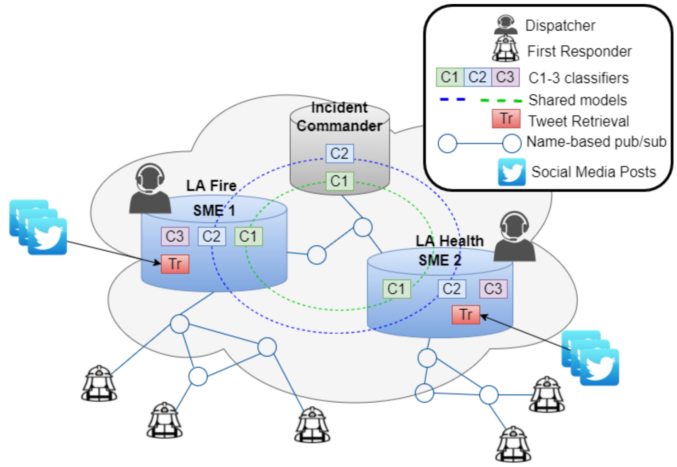

Fig. 1. Architecture of FLARE

III. Data Retrieval

We use Twitter API for fetching tweets related to a particular disaster event using keywords that relate to it as well as its type. These keywords however, can have different contextual meanings. So, by using these keywords alone, we may end up fetching a very large number of tweets that may not be relevant to the incident. A approach to explore is to fetch geo-tagged tweets in order to limit the search only to the disaster-affected areas. The difficulty is that users of many social media platforms, including Twitter, refrain from sharing their location information for privacy reasons.

So, while fetching the data using the Twitter API, we try to ensure we gather as much as possible, if not all, tweets related to the disaster while reducing the amount of unrelated tweets. It is important to not miss tweets that contain some critical information. To tackle this problem, we use following approach which has also been illustrated in Fig. 2:

Step 1: Using geo-tagged tweets to collect an initial dataset from users in the affected area:
We use the general disaster related keywords along with the *place* operator in the Twitter API to fetch tweets which are geo-tagged and fall in the disaster zone. While we performed this after the fact, one could do this based on advanced knowledge of the event (e.g., weather information). Using geo-tagging in conjunction with the general disaster related keywords helps in reducing the number of unrelated tweets by a large margin and enables us to get first-hand information from the people in the disaster zone.

Step 2: Using disaster incident specific keywords to expand the dataset:

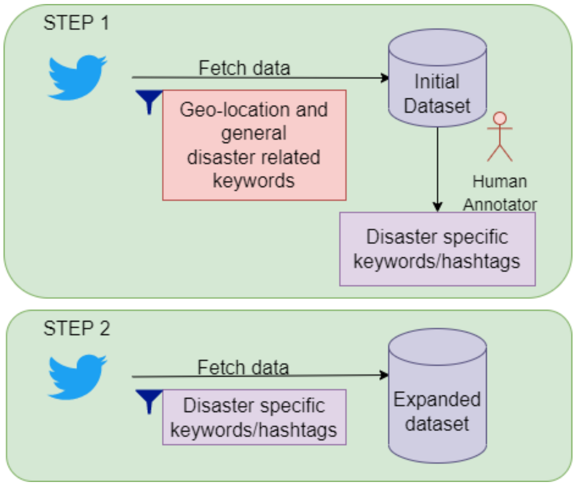

Fig. 2. Procedure for retrieval of disaster related Tweets

In this step, we leverage the information in Step 1 to get *keywords and hashtags specific to the Disaster incident*. The system uses the tweets fetched in Step 1 and extracts the list of the most frequently occurring keywords and hashtags. A human annotator may then eyeball the list and find the keywords that contains the *Disaster specific* keywords, e.g., hashtags involving the name of the disaster. Such *Disaster specific keywords and hashtags* are then used to fetch all the disaster specific tweets from the much larger category of tweets during the event's time period and its aftermath, which are not geo-tagged.

In Sec.V, we show how we are able to get most of the relevant tweets using this approach for the IDA hurricane which was the second-most damaging and intense hurricane in recent US history. Hurricane IDA lasted for a prolonged period, from Aug. 26, 2021 to Sept. 04, 2021. It affected people living in a very large area across the United States including Louisiana, Mississippi, Pennsylvania, New Jersey, New York, and Connecticut.

IV. Improved Classification Mechanism

The C1 classifier used in FLARE [1] classifies incoming tweets as either being relevant or irrelevant. However, in addition to being relevant, there are many tweets which might be related to the event but may not be *actionable*. In the paper, we have classified a number of non-actionable tweets into the *Other* category. The Other category mostly includes tweets praying and wishing for the well-being of the affected people. Since we use FLARE to propagate actionable items to the first responders, the C1 classifier should be able distinguish such tweets even though they are relevant. Such tweets are relevant to the incident but may not be the most important in assisting first responders in their task of immediately helping the victims. To address this issue upgrade the classifier proposed in FLARE to provide detailed prediction rather than just a binary classification based on relevance. The TREC-IS dataset [13] contains 70,396 distinct tweets for 75 different disaster events. Each tweet therefore belongs to one or more of 25 different classes (as shown in Table. I). The support here represents the total number of tweets that belong to the corresponding class. Since a tweet can belong to multiple classes, if we take the sum of support across all the classes, the total will exceed the count of 70,396 tweets. The incident commander in FLARE can choose which of the 24 relevant classes, contains information worth sending to first responders and should be passed on in the FLARE system for C2 classification.

TABLE I
TREC-IS LABELED DATASET CLASS DISTRIBUTION

Categories/Classes	Support	Actionable
Irrelevant	29897	
Location	28143	
MultimediaShare	23065	
ContextualInformation	4920	
Weather	6560	
Discussion	4138	
Hashtags	13997	
News	16941	
Official	3237	
EmergingThreats	7473	✓
FirstPartyObservation	1828	
Factoid	9324	
ThirdPartyObservation	15151	
MovePeople	708	✓
Sentiment	5506	
Advice	2463	
GoodsServices	115	✓
Donations	354	
ServiceAvailable	1709	✓
SearchAndRescue	61	✓
OriginalEvent	3355	
NewSubEvent	3120	✓
Volunteer	168	
CleanUp	444	
InformationWanted	300	

In Fig. 3, we show the updated C1 classifier that we use for this purpose. In this classifier model, we set the number of neurons in the output layer to 25 representing the 25 different classes in the TREC-IS dataset. We use a Sigmoid activation function in the output layer to get the probability of a tweet belonging to each one of the 25 classes to be between 0 and 1. We compare this probability against a fixed threshold to determine whether the tweet belongs to a particular class or not. Similar to FLARE, we keep the DNN model separated from the embedding layer to reduce the model size and therefore reduce the cost of model exchange and aggregation during federated learning.

V. Evaluation

A. Evaluating Data Retrieval Approach:

For the evaluation of the approach mentioned in §III, we fetched the tweets related to hurricane IDA. We used the regions affected by the hurricane and general disaster related keywords like hurricane, outage, shortage, *etc.*, to fetch the tweets. Table II shows the distribution of the tweets which

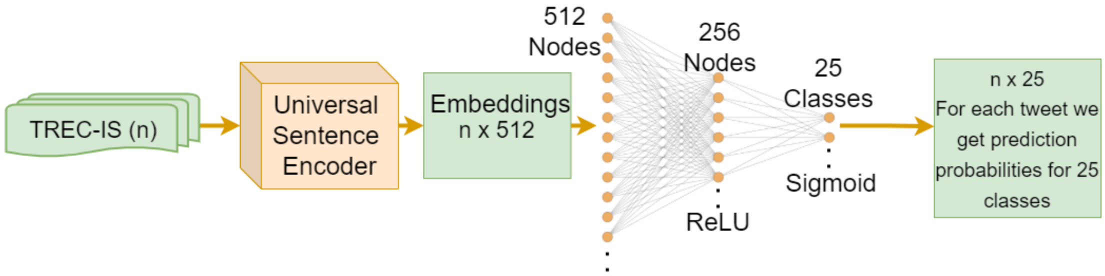

Fig. 3. Updated C1 classifier model

were geo-tagged. With this, totally we were able to collect only 9,992 tweets which was a relatively small amount. We then collected the list of most frequent keywords and hashtags and extracted the hashtags specific to IDA from this list such as: #IdaAftermath OR #HurricaneIda2021 OR #IdaRelief OR #IdaHelp. Using these hashtags, we were able to fetch 179,308 tweets that were not geo-tagged, but related to the disaster. We show the distribution of these tweets based on the date in Table III. We also made sure to exclude 'retweets' from this list as we are only interested in first-hand actionable information. So, using our data retrieval method, we were able to increase the total tweet count by 1,794.5%.

TABLE II
LOCATION-WISE GEO-TAGGED TWEET DISTRIBUTION FOR IDA HURRICANE

Location	# Tweets
Louisiana	1019
Mississippi	1059
Pennsylvania	2741
New Jersey	374
New York	4299
Connecticut	500
Total	**9992**

TABLE III
DATE-WISE DISTRIBUTION OF TWEETS WITH IDA SPECIFIC KEYWORDS/HASHTAGS

Date	# Tweets	Date	# Tweets
8/26/2021	1335	9/5/2021	2541
8/27/2021	10247	9/6/2021	2442
8/28/2021	15934	9/7/2021	2812
8/29/2021	49229	9/8/2021	2360
8/30/2021	35821	9/9/2021	2069
8/31/2021	11590	9/10/2021	1517
9/1/2021	9069	9/11/2021	649
9/2/2021	17676	9/12/2021	617
9/3/2021	7577	9/13/2021	955
9/4/2021	3873	9/14/2021	995
Total		**179308**	

B. C1 Classifier Training:

For training the model for enhanced C1 classification, we use 80% (56316 tweets) of the labeled TREC-IS dataset. The hyperparameter configuration used for training is shown in Table IV. We use 250 epochs for training each batch to ensure that the model is given the chance to train iteratively if there is scope of improvement. At the same time, we use Early Stopping, to avoid overfitting of the model. We use 10% of the test set for validation. We keep a higher value of batch size, *i.e.* 256, to ensure tweets of each batch covers most of the classes. We use the Adam optimizer [14] with loss as the metric. We got approximately the same performance with other optimizers and metrics as well.

With the above configuration, we train two different models based on different sentence embedding techniques. The first model uses the Universal Sentence Encoder [15] for embedding the tweets.The second model uses Sentence-BERT [16] which is a modification of a pretrained BERT model [17] to generate sentence embeddings.

TABLE IV
HYPERPARAMETER VALUES FOR THE MODEL

Hyperparameter	**Value/Description**
Epochs	250
Batch Size	256
Loss model	binary crossentropy
Optimizer	adam; metric = loss
Early Stopping	monitor = loss; patience = 3; min delta=0.0001; restore best weights= True

C. C1 Classifier Evaluation:

For the evaluation of the trained C1 classifier described in §IV, we compared our results on the same evaluation metrics used in TREC 2020 report, *i.e.* overall accuracy, overall F1-score and F1-score for actionable classes. Actionable classes refer to the 6 of 25 different classes in the TREC-IS classification as depicted in Table I. These actionable classes cover information related to emerging threats, search and rescue, available goods and services, *etc.*. For the evaluation, the F1-score is the harmonic mean of precision and recall. From our application's perspective high precision makes sure that the first responders do not receive a high amount of irrelevant/unrelated information. A higher recall ensures that most of the relevant information is getting classified correctly and sent to the right first responders. Thus, a higher F1-score value implies that the model is better at classifying most of relevant information without much noise in form of irrelevant/unrelated data. Our performance results are shown in Fig. 4. We use the 20% of the labeled TREC-IS dataset mentioned in Table I as the test-set for the evaluation.

Both the Universal Sentence Encoder (USE) [15] and the Sentence BERT (SBERT) [16] models have almost similar performance, but USE has a slight edge in all the three metrics used for evaluation. The metrics represented for classifiers with SBERT and USE is the average across 100 runs with different training and testing set combinations. The standard-deviation is shown in 'green' color in the Fig. 4. The low standard-deviation, across the 100 runs with different combinations of training and testing sets show that there is no data-bias while splitting the data across training and testing sets.

Since the performance with USE is better than SBERT, we submitted the classification of TREC-IS 2021-A test topics, using the USE based classifier. Fig. 5 shows the comparison of our model with the one submitted by the ucd-cs team [18].

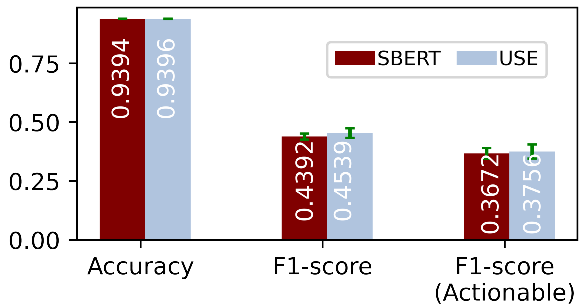

Fig. 4. Evaluation of models with SBERT and USE

That model had the best F1-score (both overall and specific to the actionable items) across all the submissions in the TREC-IS track. Here, we see that our model's F1-score (both overall and specific to actionable items) is lower, but we see a slight improvement in accuracy. Although these results indicate that our model needs to improve compared to the best-performing model, there are a number of factors of our model that are very useful for real-time classification. Firstly, since we only train the neural network in the model, without fine-tuning the sentence encoder, the model size is significantly smaller, which leads to faster model exchange and aggregation during federated learning. Secondly, the main purpose of this model is to act as a base to start the classification on any new event. Since the model can use active learning, it can be faster in adapting to new incoming data, thus improving the results over time. Our main purpose is to provide a base model which is trained sufficiently to provide a reasonable classification at the occurrence, or beginning, of a disaster. There are other useful techniques, such as ensembling and data augmentation, which can be used to further improve its performance.

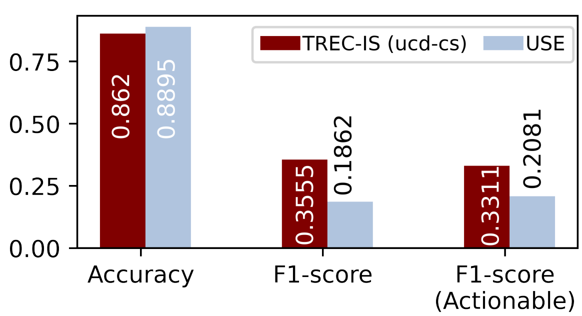

Fig. 5. USE evaluation with the best TREC-IS submission (UCD-CS)

D. Model Usage:

Subsequent to the classifier evaluation, we run the model with USE encoder on the IDA dataset to classify the tweets. As mentioned earlier, it may be up to the Incident Commander to decide which classes, out of the 24 relevant ones, he/she wants in the system for forwarding to first responders. For the classification of IDA tweets here, we select the 6 *actionable* classes as mentioned by the authors of [13]. We were only able to find 234 actionable items from the tweets which had geo-location specified in them. However, due to the keyword expansion technique mentioned in V-A, we were able to garner more tweets. By classifying these tweets, we were able to find 4,592 extra *actionable* tweets.

A major flaw of this classification model is that it compares the prediction probability for each class of a tweet against a fixed threshold of 0.5. Since the dataset we are using here is imbalanced, there are very few tweets for the actionable classes. In Table I, we see that there are only 115 and 61 tweets for the 'GoodsServices' and 'SearchAndRescue' classes. As a result of this, the classifier gives very low prediction probability for these classes. We see this impact is more evident when we run our model on the IDA dataset. As shown in Table V, none of the tweets fall under these two classes and therefore their support is 0.

TABLE V
SUPPORT OF DIFFERENT 'ACTIONABLE' CLASSES ON IDA DATASET BASED ON FIXED AND OPTIMIZED THRESHOLD VALUES

	Fixed Threshold	**Optimized Threshold**
EmergingThreats	2517	3875
MovePeople	185	3102
GoodsServices	0	9519
ServiceAvailable	1960	29822
SearchAndRescue	0	80
NewSubEvent	240	899

We make use of Youden's J statistic [3] to calculate the appropriate threshold values for each of the classes. By using these thresholds, we were able to also retrieve tweets for the minority classes from the IDA dataset. With the new thresholds, we are able to retrieve 37,295 'actionable' tweets. Table V shows the support for each class in the classification with optimized threshold values. Here, we are able to get tweets for all 6 of the 'actionable' classes.

Since we use a multilabel classifier and a single tweet can belong to multiple classes, we study the relationship between the 6 'actionable' classes and represent it using the matrix shown in Table. VI. Each cell in the matrix represents the number of common tweets between the two classes indicated by its row and column. In the diagram, we see that most of the tweets belong to 'GoodsServices' class also belong to the 'ServiceAvailable' class. This is justified as both these classes are related to services. There can also be other services which may not involve goods, and thus, we see a much higher number of tweets in 'ServiceAvailable' class as compared to 'GoodsServices'. By manually examining the tweets from different classes we find that the classifier predicts the classes fairly accurately on the new dataset. We also provide examples for various tweets and their prediction in Table VII. These examples demonstrate the ability of the classifier to efficiently extract just the relevant items from the dataset which may require assistance of the first responders.

TABLE VI
RELATIONSHIP MATRIX FOR 'ACTIONABLE' CLASSES

Classes	ET	MP	GS	SA	SAR	NSE
EmergingThreats (ET)	3875	462	10	310	0	116
MovePeople (MP)	462	3102	210	880	2	39
GoodsServices (GS)	10	210	9519	8166	59	6
ServiceAvailable (SA)	310	880	8166	29822	66	62
SearchAndRescue (SAR)	0	2	59	66	80	0
NewSubEvent (NSE)	116	39	6	62	0	899

TABLE VII
EXAMPLES OF TWEETS FROM IDA DATASET WITH PREDICTION

Tweets	Predicted Classes
#TropicalStorm #Ida forecast to strengthen...tropical storm conditions expected in the #CaymanIslands later tonight and across western #Cuba on friday…	[EmergingThreats]
Voluntary evacuation of Grand Isle, #LAWx due to #Ida with a mandatory evacuation expected midday tomorrow…	[EmergingThreats, MovePeople]
The Mayor of Grand Isle called for a voluntary evacuation this evening and a mandatory evacuation tomorrow due to Tropical Storm #IDA	[MovePeople]
Well, all ye emergency and disaster responders, trailers, and helpers on the Gulf Coast, time to saddle up. #hurricaneIda. #Ida #Louisiana ready? \n\nPS: I know you're tired.	[GoodsServices]
Hurricane #Ida is forecast to make landfall as a major hurricane along the Louisiana coast Sunday. \n\nRight now, we have hundreds of volunteers headed to help. We need hundreds more.\n\nIf you can join us, even for a few months, we urgently need you.	[GoodsServices, ServiceAvailable]
Please begin your hurricane preparations in anticipation of Hurricane Ida that is forecasted to impact the Southeast Louisiana area. Here are some hurricane safety tips from the American Red Cross #ida #haveaplan #hurricane #teamtoups #becausewelivehere too	[ServiceAvailable]
A sad update to the search for a 5-year-old boy in Lancaster County. We'll have a live report tonight at 6 on @abc27News.	[SearchAndRescue]
For those in #HurricaneIda's path. Please list others below! Be safe.	[ServiceAvailable, SearchAndRescue]
Anyone in and around Ida's path should seriously GTFO ASAP. Also, wear a damn mask because that area is surely toxic with COVID too.	[NewSubEvent]

VI. CONCLUSION

We presented FUSED, an enhancement to FLARE, adding a key component of data retrieval from social media platforms. We leverage geo-tagged posts to find incident specific keywords/hashtags, which are then used to retrieve a larger share of relevant data. We show the applicability of this approach by testing it on fetching tweets for the IDA hurricane. Using IDA specific hashtags, we were able to expand the dataset by 1,794.5%. We then update the first-level, C1 classifier in FLARE. This enables an Incident Commander to choose the sub-categories of relevant posts, he/she decides to forward to first responders. We use TREC-IS dataset to train the C1 classifier for this purpose. We further improve the overall model by selecting the optimum thresholds for the prediction probability, to deal with class imbalance. Finally, we demonstrate the applicability of these updates over the IDA hurricane dataset and show how the system can be used in real-time to classify the tweets.

In these experiments, we use the pre-trained C1 model without active and federated learning and show the strength of the base C1 classifier. The model can then use the active and federated learning features of FLARE. We plan to test these learning aspects of FLARE with the updated model next.

VII. ACKNOWLEDGEMENTS

This work was supported by the US National Science Foundation grant CNS-1818971 and the US Dept. of Commerce, NIST PSIAP award 70NANB17H188. We thank Prof. Cody Buntain for his kind guidance and promptness in providing the results for our submission to TREC-IS 2021-A test topics on very short notice.

REFERENCES

[1] V. Mittal, M. Jahanian, and K. K. Ramakrishnan, "Flare: Federated active learning assisted by naming for responding to emergencies," in *2021 IEEE 29th International Conference on Network Protocols (ICNP)*. IEEE, 2021, pp. 1–2.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.

[3] V. Bewick, L. Cheek, and J. Ball, "Statistics review 13: receiver operating characteristic curves," *Critical care*, vol. 8, no. 6, pp. 1–5, 2004.

[4] S. Utz, F. Schultz, and S. Glocka, "Crisis communication online: How medium, crisis type and emotions affected public reactions in the fukushima daiichi nuclear disaster," *Public relations review*, vol. 39, no. 1, pp. 40–46, 2013.

[5] L. Palen, "Online social media in crisis events," *Educause Quarterly.vol. 31, no. 3*, p. 76–78, 2008.

[6] M. Saunders, "Social media: California wildfires force thousands to evacuate," https://tinyurl.com/y2pfolyd, Nov. 2018.

[7] K. K. Ramakrishnan, M. Yuksel, H. Seferoglu, J. Chen, and R. A. Blalock, "Resilient Communication for First Responders in Disaster Management," in *Proceedings of 18th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, 2021.

[8] K. Zahra, M. Imran, and F. O. Ostermann, "Automatic identification of eyewitness messages on twitter during disasters," *Information processing & management*, vol. 57, no. 1, p. 102107, 2020.

[9] K. Banujan, B. T. Kumara, and I. Paik, "Twitter and online news analytics for enhancing post-natural disaster management activities," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2018, pp. 302–307.

[10] M. Y. Kabir, S. Gruzdev, and S. Madria, "Stimulate: A system for real-time information acquisition and learning for disaster management," in *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2020, pp. 186–193.

[11] B. Rachunok, C. Fan, R. Lee, R. Nateghi, and A. Mostafavi, "Is the data suitable? the comparison of keyword versus location filters in crisis informatics using twitter data," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100063, 2022.

[12] K. Lee, R. K. Ganti, M. Srivatsa, and L. Liu, "When twitter meets foursquare: tweet location prediction using foursquare," in *Proceedings of the 11th international conference on mobile and ubiquitous systems: Computing, networking and services*, 2014, pp. 198–207.

[13] C. Buntain, R. McCreadie, and I. Soboroff, "Incident streams 2020: Trecis in the time of covid-19," in *Proceedings of the Twenty-Ninth Text REtrieval Conference (TREC 2020), Gaithersburg, Maryland*, 2020.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar *et al.*, "Universal sentence encoder," *arXiv preprint arXiv:1803.11175*, 2018.

[16] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[18] C. Wang, P. Nulty, and D. Lillis, "Transformer-based multi-task learning for disaster tweet categorisation," *arXiv preprint arXiv:2110.08010*, 2021.