Mirroring Hierarchical Attention in Adversary for Crisis Task Identification: COVID-19, Hurricane Irma

Shalini Priya*
Indian Institute of Technology Patna
shalini.pcs16@iitp.ac.in

Manish Bhanu
Indian Institute of Technology Patna
manish.pcs16@iitp.ac.in

Sourav Kumar Dandapat
Indian Institute of Technology Patna
sourav@iitp.ac.in

Joydeep Chandra
Indian Institute of Technology Patna
joydeep@iitp.ac.in

ABSTRACT

A surge of instant local information on social media serves as the first alarming tone of need, supports, damage information, etc. during crisis. Identifying such signals primarily helps in reducing and suppressing the substantial impacts of the outbreak. Existing approaches rely on pre-trained models with huge historic information as well as on domain correlation. Additionally, existing models are often task specific and need auxiliary feature information. Mitigating these limitations, we introduce Mirrored Hierarchical Contextual Attention in Adversary (MHCoA2) model that is capable to operate under varying tasks of different crisis incidents. MHCoA2 provides attention by capturing contextual correlation among words to enhance task identification without relying on auxiliary information. The use of adversarial components and an additional feature extractor in MHCoA2 enhances its capability to achieve higher performance. MHCoA2 reports an improvement of $5-8\%$ in terms of standard metrics on two real-world crisis incidents over the state-of-the-art.

Keywords

Covid-19, hurricane, adversarial, hierarchical attention, support, infrastructure damage.

INTRODUCTION

During crisis, conversation over social media regarding crisis has emerged as an indispensable source to prepare automated techniques to flatten the disaster scenario consequences (Priya, S. Singh, et al. 2019, Priya, Upadhyaya, et al. 2020, Priya, Sequeira, et al. 2019). Identifying task related information such as information relevant to supports needed, available resources, infrastructure damage, casualties, etc. from social media is of utmost importance for emergency caregivers so that they can prioritize their relief activities and save lives.

However, the tweets posted during different disaster events such as Biological pandemic (COVID, Ebola, Zika, etc.) and Meteorological crisis (Hurricane, Tornado, Heat wave, etc.) have different reasons of spread and varying consequences, hence the situational tasks emerged may also be different. State-of-the-art techniques have manually designed the tasks during different disaster events (L. Li et al. 2020, Rudra, Ghosh, et al. 2015), among these tasks we select few tasks to solve using text classification approach. For this work, we consider two crisis events COVID-19 and Hurricane Irma. Lockdown during COVID-19 severely affected low-income communities, and lack of financial sources, leaving them with fewer commodities and hence more vulnerable to such disease. Such communities may have several other types of urgent needs, which can be identified through social media data. Such anticipated needs require to be explored using automated technique for the sake of the general public (the information gaps between the authority and the public, and the information need of the public). Other then “required support" there are several supports being offered and information regarding the same needs to be communicated to

*corresponding author

Table 1. Few tweets from the two datasets. These tweets are labelled relevant for the corresponding Class.

Class	Event	Example tweets
Support signals	COVID-19	i. Hi, it's Eternity0227_. I hope our donation for Wuhan Coronavirus will help people here. We donated 10000CNY to hospital ii. Nepal Government is preparing to send 2 planes to bring back Nepalese students in China who are at living in Corona affected region. iii. The head of a hospital treating #coronavirus"" Salute to this doctor. This is amazing dedication to his profession.
Infrastructure damage	Hurricane Irma	i. Photos show destruction, damage from Hurricane Maria at puerto rico. ii. Hurricane Maria damage latest: 100,000 homes without power after storm batters Caribbean iii. Hurricane Maria continues destructive path through The Caribbean sea

the deprived ones. Following (L. Li et al. 2020), tweets posted are classified into 7 categories. However, we select three major categories, i.e Type3 (donation of money, goods, and services), Type 4 (providing emotional support), and Type 5 (Help-seeking), and club them as an individual class and named them as support signals.

Other than identifying support signals during the biological pandemic, state-of-the-art has shown the importance of situational information during hurricanes, earthquakes, etc. (Imran, Alam, et al. 2020; Priya, Bhanu, et al. 2020). However, each situational information posted is extremely demanded but the information regarding infrastructure damage holds notable significance for emergency assistance (Priya, Upadhyaya, et al. 2020). Example of tweets that have been considered relevant to both tasks is shown in Table 1.

Extracting task relevant information from unstructured (brevity, presence of typos, digits, punctuation, hashtags, etc.) tweets is a challenging task. Moreover, the distribution drift between text generated from different locations (Priya, Bhanu, et al. 2020) and it's identification for different crisis tasks makes it even more challenging. Machine learning (ML) or Deep Neural Network (DNN) models trained on previous disaster events (pre-trained models) are also not suitable for the same (Priya, Upadhyaya, et al. 2020).

Limited works have been done towards classifying important information during epidemics (Ebola, plague, etc.). Such techniques rely on manually generated task conditioned features and may not perform well as the change in keyword distribution will result in the change in disaster events (Alam, Joty, et al. 2018). Among recent works (L. Li et al. 2020), has harnessed Weibo and Twitter data for exploiting natural language processing and traditional ML techniques to classify COVID-19-related information into several types of situational information. Other than COVID-19 specific data, researchers have proposed similar ML model with meta and linguistic features with other health disaster datasets (Adel and Wang 2019; Rudra, Sharma, et al. 2017). Further, word embedding approach was used for 2014 Ebola and 2016 Zika outbreaks for identifying crisis-related actionable tweets (Khatua et al. 2019).

Apart from literature's in the biological pandemic domain, similar attempts have been done in other geophysical, hydrological, and meteorological disaster events to classify situational data and used ML and DNN based models (Khatua et al. 2019; Rudra, Sharma, et al. 2017; Rudra, Ghosh, et al. 2015). They require manually engineered-features like task-specific keywords and TF-IDF vectors (Imran, Castillo, Diaz, et al. 2015) for learning the task properties. Other than the huge dependency of such techniques on pre-processing and feature engineering, adapting the model to changes accordingly is mostly infeasible when there is a change of either crisis events or tasks. Hence such approaches are time consuming and not always reliable. Additionally, embedding based techniques have also being implemented with these disasters (Basu et al. 2019). However, domain-specific tweet corpora during the early stages of outbreaks are normally scant; identifying actionable tweets during early stages is crucial for stemming the proliferation of an outbreak (Khatua et al. 2019). Moreover, these techniques require huge annotated tweet samples and crowd-sourcing to train ML or ANN models that are undesirable and time-consuming at the early stage of disaster event (Imran, Castillo, Lucas, et al. 2014). Authors (Madichetty and Sridevi 2021a) have proposed technique to identify infrastructure damage tweets using exhaustive feature sets but the features are extracted only by taking care of the dedicated task, hence can not be generalized for the health disaster scenario.

Having seen the drawbacks of pre-trained, ML and embedding based models, DNN models have an inherent capability to catch latent features, capture non-linear dependency among features automatically without any feature engineering (Nguyen et al. 2017) and is widely used for task related tweet identification in many classification tasks. Therefore, it will be helpful if we develop a uniform model that is free from manual effort of feature engineering and additional information required to capture contextual correlation. To prepare such model, we propose Mirrored Hierarchical Contextual Attention based adversarial (MHCoA2) framework which captures the contextual correlation among the text and uses them as attention to highlight relevant latent features for identifying crisis-related task tweets.

The major contributions of this paper are: (1) The proposed model uses contextual information of the input instances as attention and emphasizes the most relevant one without the need for any additional information for attention. (2) The proposed model involves a dedicated component for capturing different parts of “Task Feature space" using an adversarial strategy. (3) We apply our model for two tasks on publicly available datasets: (i)identifying support signals from COVID-19 data and, (ii)identifying infrastructure damage relevant text information from Hurricane Irma data. We choose datasets and tasks from two entirely different domains so that the robustness of MHCoA2 can be verified. We observe significant improvement in terms of standard metrics when compared with state-of-the-art techniques. Further, details are provided in the relevant sections below.

RELATED WORK

We closely examine the growing landscape of research done for all natural disasters using social media data. Existing techniques in disaster domains have seen a growth from traditional ML to deep learning (DL) techniques. Classical models use bag-of-words features to classify situational tweets generated during Earthquake and Flood, similar techniques have been used for pandemic data such as COVID-19, Ebola, and Zika, etc. (Alam, Ofli, Imran, and Aupetit 2018; Adel and Wang 2019; Ghosh et al. 2019). Popular Artificial Intelligence and Disaster Response (imran2014aidr) work use crowd-sourcing to get vast annotated samples and uni-gram, bi-gram features to train the model (Imran, Castillo, Lucas, et al. 2014). Authors (Rudra, Ghosh, et al. 2015) extract lexical and syntactic features for classifying situational information during disasters. Other ML techniques (SVM, NB, LR) consider informative keywords, intensifiers count, etc. for situational tweets identification. Although several techniques including pattern matching, language modeling, and DL (X. Li, D. Caragea, Zhang, et al. 2019; Priya, Bhanu, et al. 2018; Priya, Bhanu, et al. 2020) have been proposed, however, it is a critical task to prepare manual query and informative keywords for the success of such techniques. As the above mentioned works are highly dependent on handcrafted features, task specific keywords, and require feature engineering, it makes the task relevant tweet identification challenging, every time a disaster strikes.

Recently word-embedding techniques have received attention due to the extraction of features without manual intervention. Apart from the low availability of the current crisis data to train the models, the developed models may also suffer from distribution drift (Priya, Upadhyaya, et al. 2020), whereby the dominant feature components in the data change over time. Such techniques can not be deployed in every situation and pre-trained embeddings (Glove, ELMo, BERT, etc.) are also not helpful due to the existing large vocabulary gaps in the tweets (Basu et al. 2019; Alam, Ofli, Imran, and Aupetit 2018). Further, authors developed an IR framework by combining word embeddings and character embeddings for extracting the resource need and availability tweets during several geophysical disasters (Basu et al. 2019). Limitation is that for the same tweet keyword, embedding vector is same even if they are used in a different context. Hence, it performs poorly while there is a need to disambiguate the words in a different context. Further basic DNN models with crisis embedding are also exploited for classifying the crisis related data (Madichetty and Sridevi 2021a). Improvements using DNN’s is observed over ML with handcrafted features and embedding based techniques. All the keywords used in the tweet have been given equal weight while training CNN and MLP-CNN, but as suggested by authors (Priya, Bhanu, et al. 2020; Priya, Bhanu, et al. 2018), certain important keywords need to be given more attention to identify important signals during any disaster events. Our proposed work uses context based hierarchical attention model for identifying different task related tweets from two entirely different disaster events.

There is limited work done with the COVID-19 twitter dataset for identifying task related information.Moreover, among the literature, a vast set usually considers X-ray image data of COVID-patients for identifying such information. Similarly for identifying infrastructure damage during disasters, majority of the work consider images over text (X. Li, D. Caragea, Zhang, et al. 2018; X. Li, D. Caragea, C. Caragea, et al. 2019; Imran, Alam, et al. 2020), However, text data posted during such events are much high in numbers than images and may provide more rich set of information (like location, severity, etc. for both COVID and Hurricane).Thus processing of the textual contents holds its importance.Relevant existing state-of-the-art related to crisis domain are summarized and tabulated in Table 2.

Having seen the drawbacks of existing state-of-the-art, we aim to fill this gap by our proposed approach.

DATASET AND METHODOLOGY

This section describes the datasets, their pre-processing, and methodology.

Table 2. Summary of literature related to different task in crisis informatics domain.

Biological pandemic			
Dataset	**Objective**	**Methodology**	**Publication**
COVID-19 Twitter data	Predicting outbreak of COVID-19	Evolutionary algorithm	Schild et al. 2020
COVID-19 Weibo data	Characterisation and propagation of COVID-19 dataset	Support vector machine, Naive Baye's and Random forest	L. Li et al. 2020
COVID-19 Twitter and 4Chan data	COVID-19 related sinophobic abuse	word2vec	Jahanbin, Rahmanian, et al. 2020
COVID-19 Twitter data	information and misinformation	Text analysis and Natural language processing tools	L. Singh et al. 2020
Ebola and MERS Twitter data	Classifying epidemic information	SVM, Naive Baye's and Logistic classifier	Rudra, Sharma, et al. 2017
Ebola and MERS Twitter data	Multi-dimensional information identification	Medical database (UMLS /Metamap) and lexical features	Ghosh et al. 2019
Cholera, Famine, Refugees Twitter data	Crisis Message Classification	TF-IDF+ Lexical+ Morphological SVM, NB, RF	Adel and Wang 2019
Other pandemics			
Red River floods, Haiti earthquake, Oklahoma grass fires Twitter data	Detecting Situational awareness tweets	Bag-of-words model	**Verma2011NaturalLP**
Pakistan Earthquake Twitter data	donation, damage and casualities	n-gram features	Imran, Castillo, Lucas, et al. 2014
HDBlast, UFlood Twitter data	Extracting situational information.	Low-level lexical and syntactic features.	**rudracikm**
Nepal Earthquake Twitter data	Identification of Fact-checkable microblogs.	Bag-of-words model	**icdcn**
Nepal and Italy Earthquake Twitter data	Extracting resource needs and availabilities of tweets	Word embeddings and character embeddings	**8715653**
California Earthquake and Typhoon Hagupit Twitter data	Classification of Crisis-related data	CNN, MLP-CNN	**MADICHETTY2020962**
Nepal and Italy Earthquake Twitter data	Detection of resource tweets	CNN with manual features	Madichetty and Sridevi 2021b
Nepal Earthquake Twitter data	Identification of Infrastructure Damage Tweets	Exact-match model	Priya, Bhanu, et al. 2018
Nepal and Italy Earthquake Twitter data	Identification of Infrastructure Damage Tweets	Probabilistic approach	**priyataqe**
TREC-IS 2018 data	Identification of Actionable information	Comparison of different state-of-the-art and it's challenges	McCreadie et al. 2019

Dataset, Pre-Processing and Manual Analysis

To analyze the performance of MHCoA2, we use two publicly available datasets. These datasets are of different disasters (Biological and Meteorological) to evaluate model performance on varying objective tasks.

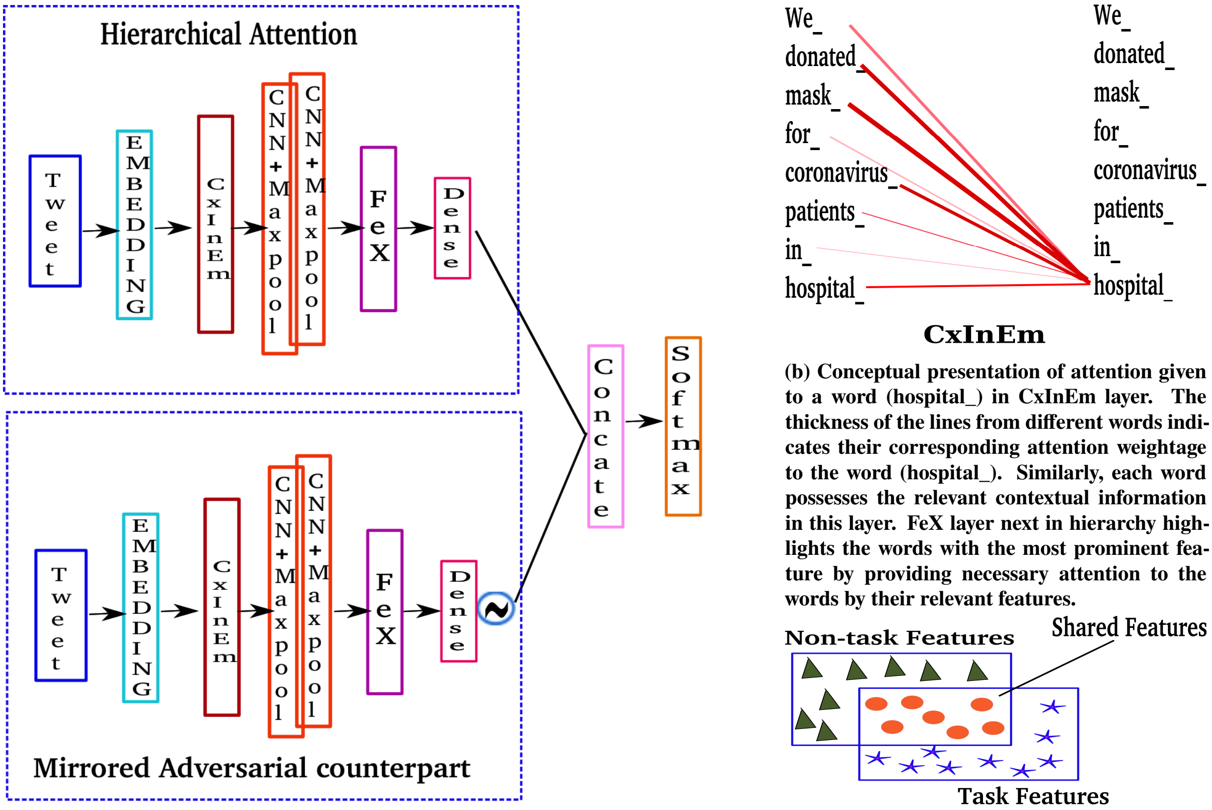
- GeoCoV19 (Qazi et al. 2020) is a large-scale Twitter dataset containing 348 million English tweets posted over a period of 90 days. However, for this work, we select a subset of English language tweets of size $0.1M$ in chronological order (as suggested in many literature Ghosh et al. 2019; Rudra, Sharma, et al. 2018; Rudra, Sharma, et al. 2017), and only tweet id and text information.
- Hurricane Irma (Alam, Ofli, and Imran 2018): Florida hurricane Irma dataset is also publicly available for use by the crisis informatics community.

Table 3. Description of publicly available datasets for the specified tasks used for the experimentation.

Events	Total tweets	Sampled Unique tweets	Relevant tweets	Irrelevant tweets
COVID-19	100,000	16731	471	500
Hurricane Irma	3,517,280	12023	907	950

Pre-processing: Available tweets contains root tweets as well as retweets and replies. We follow techniques involved in (Priya, Bhanu, et al. 2020) to remove any duplicates in both the datasets. Necessary pre-processing steps that have been done for tweet refinement. We remove non-ASCII characters, stopwords, numbers, tokens less than three characters, hyperlinks, Retweet tags and hashtags. Further we apply porter stemmer to reduce tokens to it's root words. Only refined tweets are further used for MHCoA2 model preparation.

Manual annotation: We follow standard literature while performing manual annotation for unique tweets in COVID-19 dataset (Olteanu et al. 2015; Alam, Ofli, and Imran 2018). Three professionals with good English knowledge were selected as annotators. Each annotator was first asked to identify support relevant tweets from the COVID dataset separately, i.e., without consulting each other. They were given certain instructions about the annotation and few examples related to the support signal tweets. Example tweets are shown in Table 1. Instruction sets shown below were prepared with the help of recent work on situational information classification (L. Li et al. 2020). The tweet is considered "support signals" if it reports one or more of the following:


Hierarchical Attention
Tweet
EMBEDDING
CxInEm
CNN+Maxpool
CNN+Maxpool
FeX
Dense
Concate
Softmax
Mirrored Adversarial counterpart
We_ donated_ mask_ for_ coronavirus_ patients_ in_ hospital_
We_ donated_ mask_ for_ coronavirus_ patients_ in_ hospital_
CxInEm
Non-task Features
Shared Features
Task Features

(a) Schematic view of MHCoA2. HAM (upper block) is mirrored as HAMA (lower block) in adversary. Initially, each word is given attention with contextual information in CxInEm layer. With a sequence of feature transformation by CNN, MaxPool, and another attention layer FeX. The most prominent features of words are highlighted in FeX layer. This feature transformation enhances the performance measure of the proposed model on the objective task.

(b) Conceptual presentation of attention given to a word (hospital_) in CxInEm layer. The thickness of the lines from different words indicates their corresponding attention weightage to the word (hospital_). Similarly, each word possesses the relevant contextual information in this layer. FeX layer next in hierarchy highlights the words with the most prominent feature by providing necessary attention to the words by their relevant features.

(c) Feature space schematic diagram. Though, any model can be efficient in identifying inputs with “Task Features", it might fail in situation with only “Shared Feature", which are ambiguous. Adversarial model HAMA in this scenario helps to identify such instances of “Shared Features". Ultimately, HAM (stars) and HAMA (circles) capture the whole “Task Features" space efficiently.

Figure 1. (a) Block Diagram of MHCoA2 and Operational view of each layer. (b) Concept of Contextual Attention. (c) The behavior of HAM and HAMA models and advantage of Adversarial Concept.

- If tweet contains donation information or wishes to donate materials, money, medical supplies, blood, or services for outbreak prevention and control.
- If tweet contains information regarding providing emotional support; praise or show empathy to others such as medical teams, public welfare organizations, celebrities, and the masses who supported covid-19, infected people.
- If tweet contains information regarding seeking help, medical institutions individual, etc. to seek support, needs, requests, queries, seek emotional support and/or services by volunteers or professionals.

We observed the Cohen kappa coefficient value of 84% for the COVID-19 dataset. This indicates high agreement among the annotators for COVID-19 datasets. The disagreements were resolved through mutual discussions. The manual annotation of the hurricane dataset is received upon request from the author of (Priya, Upadhyaya, et al. 2020).

It has been observed that relevant task-specific tweets are very less in number and following the standard literature's on text classification (Rudra, Sharma, et al. 2017; Rudra, Sharma, et al. 2018; Khatua et al. 2019), we maintain the balance between task relevant and irrelevant tweets while training MHCoA2 for tackling class imbalance. All details of both datasets are shown in Table 3. However, more annotated tweet samples may help in improving the performance of MHCoA2, but at the early phase of disaster events tweets generated are very less and hence the anticipated model is expected to perform well on lesser tweets.

PROPOSED MODEL: MIRRORED HIERARCHICAL CONTEXTUAL ATTENTION IN ADVERSARY

We elaborate the proposed classification model MHCoA2(ζ) in this section. The input to the model is the sequence of words (tweet tokens) $\{w_i, w_2, w_3 \dots w_{max\_length}\}$. Though we carry pre-processing, we avoid any need of explicit feature crafting from input data. The model (Figure 1a) consists of two components Hierarchical Attention Model HAM(ξ_i) and its mirrored counterpart in adversary HAMA(ξ_j). HAM contains two attention layers: CxInEm followed by FeX. For each word in the input sequence CxInEm captures a word's contextual correlation with other words however far the words are, by giving relevant weightage to every other word and assign the aggregated value as attention to the word (Figure 1b). FeX is able to highlight the most important features using softmax and feature aggregation. This hierarchical attention helps to reveal important features of the concerned task which can be observed by comparing the model with its variants having hierarchical attention removed (MHCoA2~Att). Additionally, the notion of using HAMA comes from adversarial networks (Priya, Upadhyaya, et al. 2020), working in adversary HAMA is able to generate (shared) feature vector which can not be distinguished between task or non-task features. Hence HAMA helps to tap the information which is most ambiguous (shared) in identifying task or non-task information. HAM helps to capture task features while HAMA helps to capture features that are shared between task and non-task (Figure 1c). Using these two feature vectors, MHCoA2 is capable of making an improved prediction[1] for the input sequence. The following presents the aggregated classification loss of the MHCoA2 (ζ):

$$\min_{\xi_i} \max_{\xi_j} \mathbb{E}_{x \sim P} \log(\zeta(\xi_i(x)\xi_j(x))) + \mathbb{E}_{x \sim P} \log(1 - \zeta(\xi_i(x)\xi_j(x))) \tag{1}$$

Where $x \sim P$ is word sequence from the distribution of task dataset (P). ξ is a component of main model which helps in predicting the correct label for the input instances.

Operational Steps of the Proposed Model: The model (ζ) is constructed of two components (ξ). With these components in adversary, ζ is able to precisely identify the fine features of the objective task. The following are the prediction steps of our model (ζ):

$$x_i = \xi_i(\{w_1, w_2, w_3 \dots w_{max\_length}\}) \tag{2}$$
$$x_j = \xi_j(\{w_1, w_2, w_3 \dots w_{max\_length}\}) \tag{3}$$
$$x = Concat([x_i, \sim x_j]) \tag{4}$$
$$\hat{y} = softmax(Dense(x)) \tag{5}$$

The $\hat{y} \in \{0, 1\}$ is the predicted class of the model. The ξ_j working against ξ_i, enhances ξ_i to correctly identify features of the required task. Next, we elaborate structural configuration of ξ_i model. Keras embedding is used to prepare the vector representation of token sequences which is passed as input to CxInEm layer. Apart from CxInEm and FeX layers, we used a sequence of CNN, Maxpool, Dropout layers. CNN is useful in capturing the features of neighborhood using Maxpool and dropout.The utility of CNN has already been shown in many standard literature (Madichetty and Sridevi 2021b; Nguyen et al. 2017). The operational steps of the ξ_i component are as follows:

$$e = Embedding(\{w_i, w_2, w_3 \dots w_{max\_length}\}) \tag{6}$$
$$p = CxInEm(e) \tag{7}$$
$$s = Dropout(Maxpool(CNN(p))) \tag{8}$$
$$s = Dropout(Maxpool(CNN(s))) \tag{9}$$
$$f = FeX(s) \tag{10}$$
$$x = Dense(f) \tag{11}$$

Contextual Information Embedding (CxInEm) layer is a unit of Multi-head attention (Vaswani et al. 2017), defined using 3 kernels ($W_Q, W_K, W_V \in \mathbb{R}^{d1 \times d2}$) with respective biases. $d1$, $d2$ are input-output dimensions respectively. The concept of these kernels is derived from the operation of making a query (W_Q) in database and obtaining values (W_V) based on specific key (W_K) (Vaswani et al. 2017). Analogous to this, these kernels help to provide attention based on query and key. Similarly, this layer is able to capture the relative correlation between embedded words. It infuses the features of each word with its contextual information (Figure 1b).The following equations present how each word in embedding space are imbued with the implicit features of other words in

[1] Prediction, identification, and classification words have been used interchangeably in this paper.

context:

$$q = W_Q * e + b_Q, \quad k = W_K * e + b_K, \quad v = W_V * e + b_V \tag{12}$$

$$v = \sigma(\frac{q * k^T}{\sqrt{d}})v + bias \tag{13}$$

$$p = \sigma(v) \tag{14}$$

Feature Extraction layer (FeX) It provides attention to the words by their feature values. It highlights the words with prominent features for the task:

$$g = \tanh(W * s + b) \tag{15}$$

$$f = sum(s * softmax(g), axis = 1) \tag{16}$$

The notations are having usual meaning unless specified.

EVALUATION AND PERFORMANCE MEASURES

This section describes training and deployment, and baseline techniques in different subsections.

Training & Deployment

The model is trained on 80% of data out of which 10% is used for validating the model using 10-fold cross-validation. We follow same train-validation-test split for every model (proposed and baseline models). Hyper-parameter has been selected using TPE in hyper-opt python library (Priya, Upadhyaya, et al. 2020) for every model. Training is stopped when validation loss converges. Best hyper-parameter are given as follows for MHCoA2: 100 epochs for COVID-19, 400 epochs for Hurricane, regularization used is l1, 0.5 dropout, activation: ReLu and sigmoid. Embedding used is embedding with dimensions as 128 (COVID-19), 32 (Hurricane Irma). Adam is the optimizer with accuracy as metrics and binary-cross entropy loss. We used python as programming language and few important libraries are keras, pandas, numpy, hyperopt, sklearn. All evaluation metric scores are reported on executing the model on the rest 20% of datasets. We contribute our code and annotated datasets publicly for researchers at: https://github.com/veracity-rumor2020/CRISIS.

Baseline Techniques

We compare the performance of MHCoA2 with certain standard state-of-the-art techniques and intermediate steps of MHCoA2.

- SVM, NB, and RF: Supervised learning methods such as support vector machines(SVM), naive Bayes(NB), and random forest(RF) to learn the types of unlabeled data based on n-gram features on COVID-19 dataset having very less labeled samples (L. Li et al. 2020). Moreover, similar techniques have been used by (Barnwal et al. 2019; Ghosh et al. 2019; Imran, Castillo, Lucas, et al. 2014) to classify situational tweets during different disaster events.
- Logistic Regression and XGBoost: We implement Logistic Regression(LR) and XGBoost with n-gram features of our dataset.
- C-LSTM: It combines CNN and LSTM to extract a sequence of higher-level phrase representations, and are fed into a LSTM to obtain the sentence embedding (Zhou et al. 2015) for text classification.
- CNN: This is the first DNN model to improve the performance of crisis tweet classification system in spite of exploiting manual features to train traditional ML models (Nguyen et al. 2017).
- Transformer: It is an encoder-decoder model, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely (Vaswani et al. 2017). Positional encoding and multi-head attention mechanism makes this model suitable for many recent sequence generation, machine translation and text classification tasks.
- Bidirectional Encoder Representations from Transformers (BERT): It is widely used bi-directional transformer model on variety of tasks in natural language processing. The BERT model (Devlin et al. 2018) use the transformer encoder architecture to process each token of tweet text in the full context of all tokens in forward and backward direction both.

- $CNTX \sim Adv$: This model is one of the intermediate steps where we remove the adversarial counterpart (HAMA) of our model.
- $CNTX \sim Attention$: It is an intermediate step where we do not use attention layer. This variant is important in measuring the performance of Hierarchical Attention of the proposed model.

EXPERIMENTATION, RESULT AND ANALYSIS

In this section, we investigate the performance of MHCoA2. Standard performance metrics such as Precision, Recall, F-measure, Accuracy, and ROC-curve are used for comparing the performance of MHCoA2 with different state-of-the-arts.

Performance comparison with the baselines We perform extensive experiments to show the performance of MHCoA2 compared to baseline techniques on both datasets. Table 4 shows the metrics value obtained. Table 4 reveals that among standard ML (NB, LR, SVM, RF, XGB), SVM outperforms other techniques in terms of Precision for COVID-19 dataset. Using n-gram feature, words with explicit correlation are well captured by SVM in COVID-19 but n-gram features are not enough to capture implicit contextual correlation among words (Hurricane Irma). Moreover, its Recall values for two datasets are not promising since SVM is not able to identify abundant of implicitly related word combinations. Thus resulting in a comparatively average F-measure and Accuracy. However, NB and LR models report different results from that of SVM. In COVID-19 dataset, NB and LR identifies most of the actual class but at cost of high penalty for false-positive prediction resulting in a compromised F-measure and Accuracy. A possible reason for this can be the assumption of independent relation (unable to capture implicit correlation) by these models for many word combinations in the dataset. The higher performance measures of MHCoA2 reveals without feature crafting, the model is able to capture implicit correlation among words, hence able to identify the respective task class better than these ML models.

We also record the performance of CNN and stacked C-LSTM. While observing Table 4 for COVID-19 dataset for identifying support signals, CNN is performing significantly better than C-LSTM but for infrastructure damage identification task in Hurricane data, stacked model i.e. C-LSTM performs better in terms of F-measure by a significant margin, that actually matches the performance shown by state-of-the-art (Priya, Upadhyaya, et al. 2020).

Moreover, MHCoA2 is compared with two trendy state-of-art models i.e. transformers and BERT. We record lesser F-measure and accuracy value with respect to our proposed approach, the lesser training dataset can be a possible reason for this (Madichetty, Muthukumarasamy, et al. 2021).

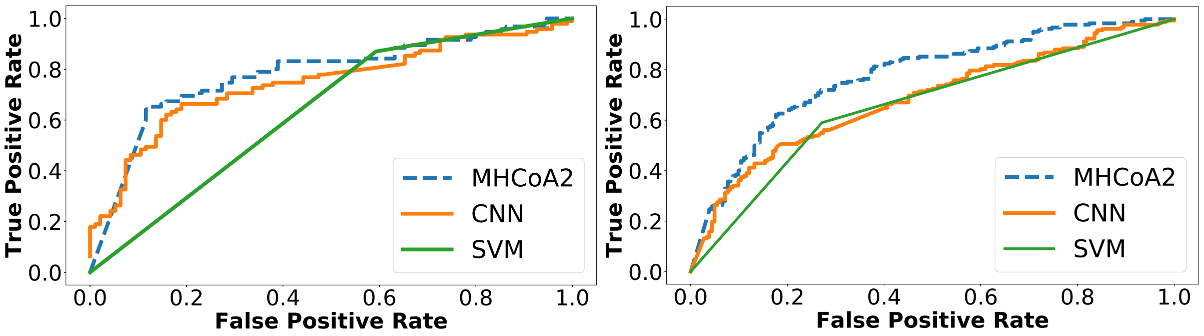
We observe MHCoA2 performs significantly better in comparison to baseline for Hurricane dataset, because of richer training dataset available in this case. We also compare the performance of our proposed approach with its variants, the results are shown in Table 4. We notice that for COVID-19, removal of adversarial layer ($MHCoA2 \sim Adv$) drops the model performance in terms of F-measure and Accuracy to lower value with respect to the removal of hierarchical attention ($MHCoA2 \sim Att$). However In case of Hurricane data we observe removal of attention ($MHCoA2 \sim Att$) drops the model performance by a huge margin, it may be due to the facts shown in multiple state-of-the-art (Priya, Bhanu, et al. 2020) that while identifying infrastructure damage tweets, there are keyword pairs that were always important for this task, so while not giving attention to the keywords, models performance drops. Additionally, the words in Hurricane dataset are having high number of implicitly correlated words that exhibit high contextual correlation which adversarial component is not able to capture better than the dedicated Hierarchical Attention part. Intuitively, this is the possible reason why most of the models' performance is better on COVID-19 than that of Hurricane dataset. The varying performance of models for different dataset reveals the fact that any baseline models are not capable to capture the task conditioned contextual information except MHCoA2. Moreover, adversarial component is able to improve identification of task features space. Contextual attention and adversarial counterpart both are necessary while we are creating a generic model for varying disaster incidents and objectives.

The performance of MHCoA2 using ROC-curve can be visualized in Figure 2. CNN is one of the standard state-of-the-art DNN method and SVM is also a better performing baseline with respect to many other, the reason we select CNN and SVM to compare the ROC-curve of MHCoA2 for both the datasets. The curve of MHCoA2 is in the top left corner for both COVID-19 and Hurricane almost everywhere, which verifies the better performance of our proposed approach.

Performance with varying embedding size We record Precision, Recall and F-measure value for COVID-19 and Hurricane dataset with varying embedding size. For both the datasets Precision and Recall value are inconsistent but maximum F-measure value is achieved at embedding size 128 for COVID-19 and 32 for Hurricane dataset as shown in Figure 3.

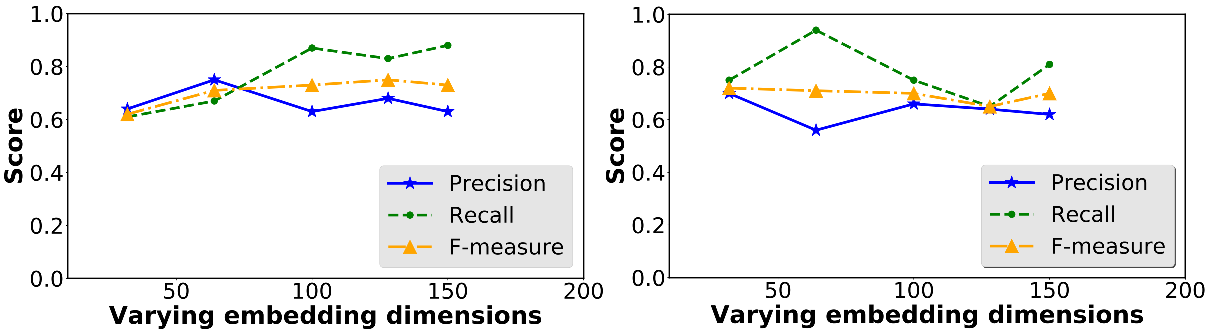
Table 4. Performance comparison of MHCoA2 with state-of-the-art and it is variants.

Method	Dataset1	Precision	Recall	F-measure	Accuracy	Dataset2	Precision	Recall	F-measure	Accuracy
State-of-the-art Techniques										
Naive Baye's (NB)	COVID-19	0.53	0.86	0.65	0.71	Hurricane Irma	0.68	0.64	0.66	0.65
Logistic Regression(LR)		0.50	0.86	0.63	0.69		0.58	0.67	0.62	0.66
Support Vector Machine (SVM)		0.86	0.61	0.71	0.65		0.59	0.68	0.63	0.66
Random Forest (RF)		0.28	0.76	0.40	0.63		0.41	0.69	0.51	0.62
XG Boost (XGB)		0.15	0.72	0.24	0.53		0.31	0.63	0.42	0.57
CNN		0.78	0.65	0.71	0.73		0.73	0.46	0.57	0.65
C-LSTM		0.64	0.69	0.67	0.62		0.69	0.62	0.65	0.67
Transformer		0.46	0.43	0.44	0.50		0.63	0.48	0.55	0.60
BERT		0.61	0.47	0.53	0.56		0.59	0.55	0.56	0.60
MHCoA2 & Variants										
MHCoA2~Adv	COVID-19	0.63	0.77	0.70	0.66	Hurricane Irma	0.64	0.70	0.67	0.66
MHCoA2~Att		0.64	0.81	0.72	0.68		0.67	0.57	0.62	0.64
MHCoA2		0.67	0.83	**0.75**	**0.79**		0.70	0.75	**0.72**	**0.71**



(a) ROC on COVID-19 data **(b) ROC on Hurricane Irma data**

Figure 2. ROC-curve of MHCoA2, CNN and SVM on (a) COVID-19 and (b) Hurricane Irma



(a) Performance measure on COVID-19 data **(b) Performance measure on Hurricane Irma data**

Figure 3. The performance measure of MHCoA2 on varying embedding dimension.

Failure Analysis and Discussion

Example tweets indexed as ⋆ in Table 5 are the tweets that are only correctly identified by MHCoA2 but other techniques failed. The colored tokens are uncommon in the dataset but contextually similar to the terms used to represent support and damage, hence our technique identifies such tweets but other techniques fail. Though MHCoA2 performs better than existing baselines, there were certain tweets where all techniques including proposed one failed and couldn't identify task relevant tweets. We list few such tweets in Table 5. First three ⊙ indexed examples are from COVID-19 dataset and next two are from Hurricane Irma dataset. We searched the dataset to find the frequency of the unigrams and bigrams in which the highlighted keywords appear. We observed that the keywords, "extend remote working", "Real Estate", "grocery chain", do not appear in any other tweets. The keyword “diagnstic msk”, “riped" and “swpt" occurs only once and is misspelled. Thus we assume that tweets having rarely used tokens and tweets containing misspelled tokens are supposed to be incorrectly labeled by any method. However, by increasing training dataset of any crisis events, such problems can be possibly addressed.

Table 5. Few examples of failure tweets. For better readability we keep original tweet representation in the Table.

⋆*Heartbreaking*! *A* 9 − *year* − *old girl brings dumplings to her nurse mother*, *who has been fighting* frontline against the #coronavirus, a ;time when families are supposed to be together. #RealHero
⋆*Hurricane Maria hammers Puerto Rico*, *causes wide destruction*
⊙BREAKING: From Feb 3, #HongKong gov’t departments extend remote working arrangements for staff. The arrangement will reâ
⊙his shouldnt be happening to people. 1 the diagnstic msk has to be available world wide before fly more people out. You cant just shut China up and not give them any help!
⊙cetain grocery chain still recovering damge from hurricane irma.
⊙the ground did not shake to alert people before the waves riped buildings from their foundations and swpt

MHCoA2 captures the contextual similarity better due to hierarchical attention layers but presently we have used it with keras embedding only, so it can be a possible limitation but we plan to add some advanced embeddings for the same. Moreover, as MHCoA2 is free from task specific feature engineering, MHCoA2 can be scaled for similar other text classification tasks where the relevant tweets (positive class) available dataset is very less.

CONCLUSION

In this paper, we propose MHCoA2, a mirrored hierarchical attention based mechanism for identifying different task relevant tweets generated during different disaster events. Major goal of MHCoA2 is to identify support signal tweets from COVID-19 dataset and infrastructure damage tweets from Hurricane Irma dataset by giving importance to the keywords responsible for task identification. We address major challenges of unstructured text available, contextual variation in text of different tasks, and efforts required for feature engineering. We compare our proposed approach with several state-of-the-art that includes standard ML and DNN models. We observe a relative improvement ranging from 5 − 8% against best performing baseline in F-measure and Accuracy for both COVID and Hurricane dataset. Further, we check MHCoA2 performance with its two variants as well to show the need for these components in the model. Investigation reveals that none of the state-of-the-art or variant is sufficient to be readily applied when there is a change in dataset with varying objectives. In future studies, we aim to apply the proposed approach with added task and image information as well.

REFERENCES

Adel, G. and Wang, Y. (2019). “Arabic twitter corpus for crisis response messages classification”. In: *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 498–503.

Alam, F., Joty, S., and Imran, M. (2018). “Domain adaptation with adversarial training and graph embeddings”. In: *arXiv preprint arXiv:1805.05151*.

Alam, F., Ofli, F., and Imran, M. (2018). “Crisismmd: Multimodal twitter datasets from natural disasters”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 12. 1.

Alam, F., Ofli, F., Imran, M., and Aupetit, M. (2018). “A twitter tale of three hurricanes: Harvey, irma, and maria”. In: *arXiv preprint arXiv:1805.05144*.

Barnwal, D., Ghelani, S., Krishna, R., Basu, M., and Ghosh, S. (2019). “Identifying fact-checkable microblogs during disasters: a classification-ranking approach”. In: *Proceedings of the 20th International Conference on Distributed Computing and Networking*, pp. 389–392.

Basu, M., Shandilya, A., Khosla, P., Ghosh, K., and Ghosh, S. (2019). “Extracting resource needs and availabilities from microblogs for aiding post-disaster relief operations”. In: *IEEE Transactions on Computational Social Systems* 6.3, pp. 604–618.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805*.

Ghosh, S., Rudra, K., Ghosh, S., Ganguly, N., Podder, S., Balani, N., and Dubash, N. (2019). “Identifying multi-dimensional information from microblogs during epidemics”. In: *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 224–230.

Imran, M., Alam, F., Qazi, U., Peterson, S., and Ofli, F. (2020). "Rapid damage assessment using social media images by combining human and machine intelligence". In: *arXiv preprint arXiv:2004.06675*.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S. (2015). "Processing social media messages in mass emergency: A survey". In: *ACM Computing Surveys (CSUR)* 47.4, pp. 1–38.

Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. (2014). "AIDR: Artificial intelligence for disaster response". In: *Proceedings of the 23rd international conference on world wide web*, pp. 159–162.

Jahanbin, K., Rahmanian, V., et al. (2020). "Using Twitter and web news mining to predict COVID-19 outbreak". In: *Asian Pacific Journal of Tropical Medicine* 13.8, p. 378.

Khatua, A., Khatua, A., and Cambria, E. (2019). "A tale of two epidemics: Contextual Word2Vec for classifying twitter streams during outbreaks". In: *Information Processing & Management* 56.1, pp. 247–257.

Li, L., Zhang, Q., Wang, X., Zhang, J., Wang, T., Gao, T.-L., Duan, W., Tsoi, K. K.-f., and Wang, F.-Y. (2020). "Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo". In: *IEEE Transactions on Computational Social Systems* 7.2, pp. 556–562.

Li, X., Caragea, D., Caragea, C., Imran, M., and Ofli, F. (2019). "Identifying disaster damage images using a domain adaptation approach". In: *Proceedings of the 16th International Conference on Information Systems for Crisis Response And Management*.

Li, X., Caragea, D., Zhang, H., and Imran, M. (2018). "Localizing and quantifying damage in social media images". In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 194–201.

Li, X., Caragea, D., Zhang, H., and Imran, M. (2019). "Localizing and quantifying infrastructure damage using class activation mapping approaches". In: *Social Network Analysis and Mining* 9.1, pp. 1–15.

Madichetty, S., Muthukumarasamy, S., and Jayadev, P. (2021). "Multi-modal classification of Twitter data during disasters for humanitarian response". In: *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15.

Madichetty, S. and Sridevi, M. (2021a). "A novel method for identifying the damage assessment tweets during disaster". In: *Future Generation Computer Systems* 116, pp. 440–454.

Madichetty, S. and Sridevi, M. (2021b). "A stacked convolutional neural network for detecting the resource tweets during a disaster". In: *Multimedia tools and applications* 80.3, pp. 3927–3949.

McCreadie, R., Buntain, C., and Soboroff, I. (2019). "Trec incident streams: Finding actionable information on social media". In:

Nguyen, D., Al Mannai, K. A., Joty, S., Sajjad, H., Imran, M., and Mitra, P. (2017). "Robust classification of crisis-related data on social networks using convolutional neural networks". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 11. 1.

Olteanu, A., Vieweg, S., and Castillo, C. (2015). "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*, pp. 994–1009.

Priya, S., Bhanu, M., Dandapat, S. K., Ghosh, K., and Chandra, J. (2018). "Characterizing infrastructure damage after earthquake: A split-query based ir approach". In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 202–209.

Priya, S., Bhanu, M., Dandapat, S. K., Ghosh, K., and Chandra, J. (2020). "TAQE: tweet retrieval-based infrastructure damage assessment during disasters". In: *IEEE transactions on computational social systems* 7.2, pp. 389–403.

Priya, S., Sequeira, R., Chandra, J., and Dandapat, S. K. (2019). "Where should one get news updates: Twitter or Reddit". In: *Online Social Networks and Media* 9, pp. 17–29.

Priya, S., Singh, S., Dandapat, S. K., Ghosh, K., and Chandra, J. (2019). "Identifying infrastructure damage during earthquake using deep active learning". In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 551–552.

Priya, S., Upadhyaya, A., Bhanu, M., Kumar Dandapat, S., and Chandra, J. (2020). "EnDeA: Ensemble based Decoupled Adversarial Learning for Identifying Infrastructure Damage during Disasters". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 1245–1254.

Qazi, U., Imran, M., and Ofli, F. (2020). "GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information". In: *SIGSPATIAL Special* 12.1, pp. 6–15.

Rudra, K., Ghosh, S., Ganguly, N., Goyal, P., and Ghosh, S. (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*, pp. 583–592.

Rudra, K., Sharma, A., Ganguly, N., and Imran, M. (2017). "Classifying information from microblogs during epidemics". In: *Proceedings of the 2017 international conference on digital health*, pp. 104–108.

Rudra, K., Sharma, A., Ganguly, N., and Imran, M. (2018). "Classifying and summarizing information from microblogs during epidemics". In: *Information Systems Frontiers* 20.5, pp. 933–948.

Schild, L., Ling, C., Blackburn, J., Stringhini, G., Zhang, Y., and Zannettou, S. (2020). "" go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19". In: *arXiv preprint arXiv:2004.04046*.

Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., and Wang, Y. (2020). "A first look at COVID-19 information and misinformation sharing on Twitter". In: *arXiv preprint arXiv:2003.13907*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). "Attention is all you need". In: *Advances in neural information processing systems*, pp. 5998–6008.

Zhou, C., Sun, C., Liu, Z., and Lau, F. (2015). "A C-LSTM neural network for text classification". In: *arXiv preprint arXiv:1511.08630*.