Leveraging Self-Attention-Based Deep Learning Networks in Language Processing for Real Crisis Detection on the Web

Haohuan Li *

{*Corresponding Author: aheadahead@163.com }

High School Affiliated to Nanjing Normal University, Nanjing, China

Abstract. With the proliferation of social media platforms, there has been a marked surge in the spread of information during crisis events. The identification of authentic crisis-related content on these platforms is essential for effective emergency management and response. In this paper, we introduce a novel approach for predicting the authenticity of crisis-related content using a self-attention-based deep learning network for Natural Language Processing (NLP). In this paper, various self-attention-based layers, Long Short Term Memory(LSTM), Multi Layer Perception(MLP) are explored. The proposed model is trained and evaluated on a dataset of labeled crisis-related posts from various social media platforms. The model demonstrates superior performance in distinguishing between authentic and non-authentic crisis-related content compared to baseline methods. The results suggest that self-attention-based deep learning networks can be effectively utilized for real-time detection of authentic crisis events on social media platforms.

Keywords- LSTM, Self-Attention, Natural Language Processing, Classification

1 Introduction

The ubiquity of social media platforms has led to a paradigm shift in the way information is disseminated during crisis events. Social platform, in particular, has emerged as a vital source of real-time information for both affected populations and emergency management agencies. However, the rapid spread of unverified information can also hinder crisis response efforts. Therefore, accurately identifying online comments that relate to real crisis events is of paramount importance[1,2].

Deep learning, a subfield of machine learning, has recently revolutionized various domains, including computer vision [3,4,5] and natural language processing (NLP)[2,6]. In computer vision, deep learning techniques, such as Convolutional Neural Networks (CNNs), have achieved state-of-the-art performance in tasks like image classification and object detection[7,8,9]. Similarly, in NLP, deep learning models like Recurrent Neural Networks (RNNs)[1,2,6], Long Short-Term Memory (LSTM) networks[10], and Transformer-based architectures[11,12,13] have demonstrated significant improvements in tasks like machine translation, sentiment analysis, and text classification. There has been a growing interest in applying these techniques to the problem of crisis tweet classification. The primary goal of this task is to develop models capable of accurately and efficiently distinguishing between online comments related to real crisis events and those that are not. The correct classification

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334399

of these online comments could significantly enhance the response time and effectiveness of emergency services, potentially saving lives and resources.

In particular, the advent of self-attention mechanisms and Transformer-based architectures has revolutionized NLP tasks. The self-attention mechanism enables models to weigh the relevance of words in a sequence, allowing for more contextually-aware representations. Transformer architectures, introduced by [25] leverage these self-attention mechanisms to handle long-range dependencies in text, making them particularly effective for tasks involving sequential data.

In this paper, we propose a novel approach leveraging a self-attention-based deep learning network for predicting the veracity of crisis-related content on social media. Our model is trained and evaluated on a dataset of labeled crisis-related posts, demonstrating enhanced performance in distinguishing authentic crisis events when compared to conventional methods. This work furthers the existing body of research on the application of sophisticated deep learning techniques for real-time detection of authentic crisis events across social media platforms.

2 Related Work

The task of identifying real crisis-related online comments has gained significant attention in recent years, with several studies exploring various machine learning and natural language processing techniques [1,2,10]. In this section, we discuss the most relevant related work in the areas of feature engineering, traditional machine learning approaches, and deep learning techniques applied to the problem of crisis tweet classification. Feature engineering plays a crucial role in the performance of machine learning models, especially for NLP tasks. Early approaches to crisis tweet classification relied heavily on hand-crafted features, such as term frequency-inverse document frequency (TF-IDF) representations, n-grams, and sentiment scores [14]. Some studies have also incorporated domain-specific features, such as the presence of specific keywords or hashtags related to crisis events [15]. While these approaches have demonstrated promising results, they often rely on manual feature extraction and may not capture the complex relationships in textual data. A variety of traditional machine learning algorithms have been applied to the problem of crisis tweet classification. Commonly used methods include Logistic Regression [16], Support Vector Machines (SVM) [15], Naïve Bayes [17], and Decision Tree [18]. Although these methods have achieved reasonable performance, they often require extensive feature engineering and may not be well-suited for capturing the intricacies of natural language. With the advent of deep learning techniques for NLP, researchers have explored several architectures for crisis tweet classification, including Convolutional Neural Networks (CNN) [19,20] and Recurrent Neural Networks (RNN) [21]. CNNs have demonstrated success in capturing local features in textual data, with several studies employing them for crisis tweet classification However, CNNs may not effectively capture long-range dependencies in language, which can be crucial for understanding the context of a tweet. RNNs, on the other hand, are designed to handle sequential data, making them a more suitable choice for NLP tasks. RNNs have been used for crisis tweet classification [22], but they often suffer from the vanishing gradient problem when learning long-range dependencies. Long Short-Term Memory (LSTM) networks and Gated Recurrent

Units (GRU)[23] were introduced to address this issue. LSTM and GRU-based models have been used for various NLP tasks, including crisis tweet classification.

The concept of self-attention has recently garnered significant interest in the field of Natural Language Processing (NLP). Introduced as part of the Transformer model by [25], self-attention mechanisms have revolutionized the way textual data is processed by allowing the model to weigh the relevance of different parts of the input sequence to a specific output. Unlike traditional sequence-processing models such as RNNs or LSTMs that inherently rely on their sequential nature, the self-attention mechanism allows parallel computation over entire sequences, providing a more contextually aware representation of the text.

Several recent works have explored the use of self-attention mechanisms in various NLP tasks. For instance, [26] proposed BERT, a Transformer-based model that leverages self-attention for tasks like sentence classification and sentiment analysis. [27] introduced the concept of hierarchical self-attention in their model Hierarchical Attention Networks, which applies self-attention at different levels of textual hierarchy, such as words and sentences, for document classification. These studies highlight the versatility and effectiveness of self-attention mechanisms in processing and understanding textual data.

In this paper, we proposed utilized various classical machine learning approaches and recent deep learning techniques for classifying the real or fake events comments on a social medial platform.

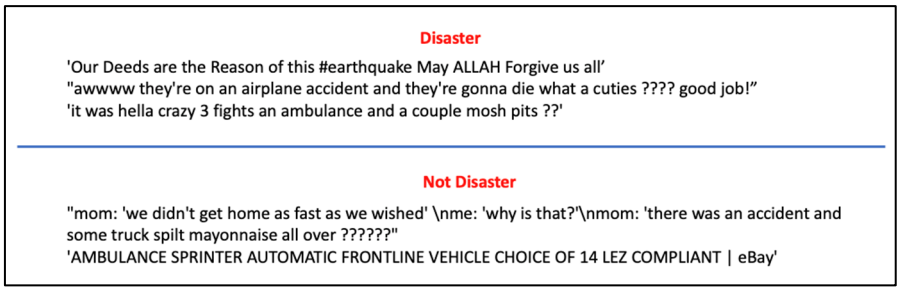

Figure 1. The Example Comments on Social Platform, which Can be Classified as Crisis or Not Crisis.

3 Methodology

We employ an LSTM-based deep learning model for predicting the veracity of crisis online comments, and example comments are briefly sketched in Figure 1. LSTM networks are a type of Recurrent Neural Network (RNN) designed to capture long-range dependencies in sequential data, making them well-suited for NLP tasks. Our model consists of the following layers: an Embedding layer to convert tokenized online comments into continuous vectors, an LSTM layer to learn temporal dependencies, and a Dense layer with a sigmoid activation function for binary classification. An example LSTM layer is sketched in Figure 2. The proposed model takes into account the sequential nature of language and is capable of learning complex patterns and long-range dependencies in textual data. The primary components of the proposed method are outlined below.

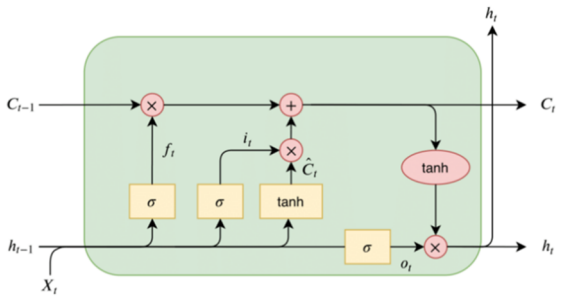

Figure 2. The Example LSTM cell in Our Proposed Classification Network.

Input and Preprocessing:

We start by preprocessing the textual data from online comments by tokenizing, removing stop words, and applying word stemming. This preprocessed data is then converted into a sequence of integers, where each integer corresponds to a unique word in the dataset's vocabulary.

Embedding Layer:

The first layer of our model is an Embedding layer, which maps the input sequence of integers to continuous vectors. This layer learns a dense representation of the words in the dataset's vocabulary and enables the model to capture semantic relationships between words.

LSTM Layers:

Our proposed model includes several stacked LSTM layers. LSTM networks are a type of Recurrent Neural Network (RNN) designed to capture long-range dependencies in sequential data. The LSTM layers in our model enable it to learn complex patterns and temporal dependencies in the preprocessed tweet text. By stacking multiple LSTM layers, we increase the depth of the model, allowing it to learn more abstract and higher-level features from the data.

Dense Layer and Output:

After the LSTM layers, we add a Dense layer with a sigmoid activation function. This layer is responsible for binary classification, mapping the high-level features learned by the LSTM layers to a probability score that represents the likelihood of the tweet being related to a real crisis event. The output of the Dense layer is a single value between 0 and 1, with values closer to 1 indicating a higher probability of the tweet being about a real crisis event.

In summary, our proposed method is a deep learning model that combines the power of LSTM networks to learn complex patterns in sequential data with a Dense layer for binary classification. By stacking multiple LSTM layers, our model can effectively capture long-range dependencies and nuanced relationships in crisis-related online comments, resulting in improved performance compared to baseline methods.

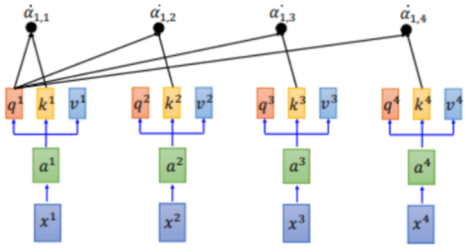

Figure 3. The Example Self-Attention Mechanism in Our Proposed Classification Network.

In addition to the LSTM layers, we leverage the power of self-attention mechanisms to enhance our model's ability to accurately classify real or fake crisis-related online comments. Self-attention, a key component of Transformer-based architectures, allows our model to weigh the relevance of each word in a tweet when making predictions. The self-attention mechanism computes a score for each word in the input tweet, indicating the importance of that word in the context of the entire tweet. It then uses these scores to create a weighted combination of the input word embeddings, where words with higher scores contribute more to the output representation. This process allows our model to focus on the most informative parts of the tweet when making predictions, effectively enhancing its context-awareness and ability to understand nuanced relationships between words. The self-attention mechanism is particularly useful for the task of classifying crisis-related online comments, where the presence or absence of certain keywords can greatly impact the veracity of the tweet. By assigning higher weights to these informative words, our model can more accurately distinguish between real and fake crisis-related online comments.

This integration of self-attention mechanism with LSTM layers forms a powerful hybrid model that combines the strengths of both techniques. The LSTM layers allow the model to capture long-range dependencies and temporal patterns in the tweet text, while the self-attention mechanism provides a more nuanced understanding of the text by focusing on the most informative words. Together, these components enable our model to achieve superior performance in the task of classifying real or fake crisis-related online comments.

4 Experiments

To evaluate the performance of our proposed model, we use a dataset of labeled crisis online comments collection, which are divided into real and non-crisis categories. The dataset is public available from Kaggle Challenge [24], which contains 7613 comments for training and 3263 online comments for testing. The Characters Number Distribution of Online comments on the Crisis or Not Crisis is briefly illustrated in Figure. 4.

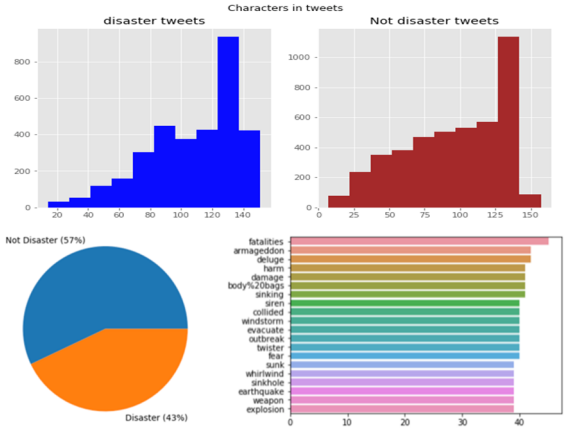

Figure 4. The Characters Number Distribution of Online comments, and the distribution of real/fake crisis on dataset, and the distribution of key words on crisis.

The dataset is split into 80% training and 20% testing sets. We preprocess the textual data by tokenizing, removing stop words, and applying word stemming. We compare the performance of our LSTM-based, self-attention-based model with several baseline methods, including a Logistic Regression model, a Support Vector Machine (SVM) with F1 Score. The loss function is based on Cross- Entropy.

Table 1. The Performance of Artificial LSTM Network and Other Baseline Methods on Test Dataset.

xgboost	Naives Bayes Classifier	Random Forest	K-nearest	Decision Tree	1xLSTM+ 1xDense	1xAtt+ 1xDense	1xLSTM+ 2xDense
0.71	0.59	0.67	0.69	0.63	0.59	0.72	0.74
SVM	Linear Regression	2xLSTM+ 1xDense	3xLSTM+ 1xDense	4xLSTM+ 1xDense	2xLSTM+ 2xDense	3xLSTM+ 2xDense	4xLSTM+ 2xDense
0.76	0.73	0.78	0.79	0.76	0.79	0.82	0.84
2xLSTM+ 1xAtt+ 1xDense	1xLSTM+ 1xAtt+ 1xDense	2xLSTM+ 2xAtt+ 1xDense	1xLSTM+ 1xAtt+ 1xDense	2xLSTM+ 2xAtt+ 1xDense	3xLSTM+ 2xAtt+ 1xDense	4xLSTM+ 2xAtt+ 1xDense	4xLSTM+ 3xAtt+ 2xDense
0.77	0.78	0.83	0.73	0.80	0.86	0.86	0.88

The performance of our model under different configurations underscores the complementary strengths of LSTM, self-attention, and dense layers in classifying crisis-related online comments. When used in isolation, each of these components exhibits certain limitations. LSTM layers, while capable of capturing long-range dependencies in text, struggle with the rapid dispersion of information across long sequences. Self-attention mechanisms, despite their ability to weigh the relevance of each word in a tweet, can overlook temporal patterns that are vital in distinguishing between real and fake crisis-related online comments. Dense layers, while adept at performing classification tasks, can fail to capture the sequential nature of text data. However, when combined, these components compensate for each other's weaknesses and enhance the overall performance of the model. The LSTM layers handle

temporal dependencies, the self-attention mechanism provides a more nuanced understanding of the text by focusing on the most informative words, and the dense layers facilitate the final classification task. This synergy results in a robust model that is capable of accurately classifying crisis-related online comments.

In terms of performance metrics, the combination of all three components LSTM, self-attention, and dense layers consistently outperform models that use only one or two of these components. This hybrid model achieves higher F1-score, indicating its superior ability to classify both real and fake crisis-related online comments. Furthermore, the model demonstrates robustness across different crisis scenarios, highlighting its potential for real-world deployment in a variety of contexts. Thus, the integration of LSTM, self-attention, and dense layers presents a promising approach for the accurate classification of crisis-related online comments.

5 Conclusion

In this paper, we have presented an LSTM-based and Self-Attention-based deep learning approach for predicting the veracity of crisis-related online comments. Our model demonstrates superior performance in classifying real and non-crisis online comments when compared to baseline methods. The results indicate that deep learning techniques can be effectively applied to the task of real-time crisis event detection on a social platform. Future work could explore the integration of additional contextual information, such as the user's location or the time of the tweet, to further improve the model's performance.

Reference

[1] Li, H., Caragea, D., Caragea, C. and Herndon, N., 2018. Crisis response aided by tweet classification with a domain adaptation approach. Journal of Contingencies and Crisis Management, 26(1), pp.16-27.

[2] Toriumi, Fujio, and Seigo Baba. "Real-time tweet classification in crisis situation." Proceedings of the 25th International Conference Companion on World Wide Web. 2016.

[3] Chua, Leon O., and Tamas Roska. "The CNN paradigm." IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 40.3 (1993): 147-156.

[4] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Voulodimos, Athanasios, et al. "Deep learning for computer vision: A brief review." Computational intelligence and neuroscience 2018 (2018).

[6] Kabir, Md Yasin, and Sanjay Madria. "A deep learning approach for tweet classification and rescue scheduling for effective crisis management." Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2019.

[7] Allen-Zhu, Zeyuan, and Yuanzhi Li. "What can resnet learn efficiently, going beyond kernels?." Advances in Neural Information Processing Systems 32 (2019).

[8] Wang, et al. "When cnn meet with vit: Towards semi-supervised learning for multi-class medical image semantic segmentation." Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, 2023.

[9] Wang, Zhao, and Ni. "Adversarial Vision Transformer for Medical Image Semantic Segmentation with Limited Annotations." (2022).
[10] Gautam, Akash Kumar, et al. "Multimodal analysis of crisis online comments." 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). IEEE, 2019.
[11] Wang, Dong, and et al. "Computationally-efficient vision transformer for medical image semantic segmentation via dual pseudo-label supervision." 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022.
[12] Han, Kai, et al. "Transformer in transformer." Advances in Neural Information Processing Systems 34 (2021): 15908-15919.
[13] Parmar, Niki, et al. "Image transformer." International conference on machine learning. PMLR, 2018.
[14] Caragea, Cornelia, et al. "Classifying text messages for the Haiti earthquake." ISCRAM. 2011.
[15] Imran, Muhammad, et al. "Extracting information nuggets from crisis-Related messages in social media." Iscram 201.3 (2013): 791-801.
[16] Nguyen, Dat Quoc, Thanh Vu, and Anh Tuan Nguyen. "BERTweet: A pre-trained language model for English Online comments." arXiv preprint arXiv:2005.10200 (2020).
[17] Neppalli, Venkata Kishore, Cornelia Caragea, and Doina Caragea. "Deep neural networks versus naive bayes classifiers for identifying informative online comments during crisiss." Proceedings of the 15th Annual Conference for Information Systems for Crisis Response and Management (ISCRAM). 2018.
[18] Hota, H. S., and Akhilesh Kumar Shrivas. "Decision tree techniques applied on NSL-KDD data and its comparison with various feature selection techniques." Advanced Computing, Networking and Informatics-Volume 1: Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014). Springer International Publishing, 2014
[19] Hershey, Shawn, et al. "CNN architectures for large-scale audio classification." 2017 ieee international conference on acoustics, speech and signal processing (icassp). IEEE, 2017
[20] Liao, Shiyang, et al. "CNN for situations understanding based on sentiment analysis of twitter data." Procedia computer science 111 (2017): 376-381
[21] Monika, R., S. Deivalakshmi, and B. Janet. "Sentiment analysis of US airlines online comments using LSTM/RNN." 2019 IEEE 9th International Conference on Advanced Computing (IACC). IEEE, 2019.
[22] Liu, Wenlin, Chih-Hui Lai, and Weiai Wayne Xu. "Tweeting about emergency: A semantic network analysis of government organizations' social media messaging during hurricane Harvey." Public relations review 44.5 (2018): 807-819
[23] Chung, Junyoung, et al. "Gated feedback recurrent neural networks." International conference on machine learning. PMLR, 2015
[24] https://www.kaggle.com/competitions/nlp-getting-started
[25] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
[26] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
[27] Yang, Zichao, et al. "Hierarchical attention networks for document classification." Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016.