Preprint accepted at the 14th International Conference on Computer and Knowledge Engineering (ICCKE 2024). Find the published version at https://ieeexplore.ieee.org/document/10874505

Supervised Contrastive Learning for Short Text Classification in Natural Language Processing

Mitra Esmaeili
Perlab, Faculty of Electrical and Computer Engineering
University of Birjand
Birjand, Iran
mitra.esmaeili@birjand.ac.ir

Hamed Vahdat-Nejad*
Perlab, Faculty of Electrical and Computer Engineering
University of Birjand
Birjand, Iran
vahdatnejad@birjand.ac.ir

Leila Rabiei
Iran Telecommunication Research Center (ITRC)
Tehran, Iran
l.rabiei@itrc.ac.ir

***Abstract*— In recent years, the swift progress in information retrieval technologies has positioned text classification as a key area of research. Classifying short texts represents a major challenge within natural language processing. Given the growing prevalence of social media during critical events like hurricanes, accurately categorizing these texts is essential for facilitating relief operations. Tweets are concise and have an informal tone, which creates unique challenges for effective classification. Supervised contrastive learning has recently become popular as a strong machine learning method. It offers significant improvements over traditional approaches, especially in the area of natural language processing. This paper introduces a supervised contrastive learning methodology designed to enhance the accuracy of short-text classification while maintaining the model's generalization capability. Our approach consistently surpasses existing state-of-the-art techniques, delivering better accuracy and stability across different ranges of text classification tasks.**

Keywords*— *Supervised contrastive learning; Text classification; Natural language processing; Semantic understanding

I. INTRODUCTION

Natural Language Processing (NLP) is a vital area within artificial intelligence dedicated to comprehending and interpreting human language [1]. A fundamental task within this domain is text classification, which involves assigning specific labels to input texts [2, 3]. The rapid growth of social media and the resulting surge in short textual data have created an urgent need for effective classification methods. One significant application is the classification of tweets related to crises and natural disasters, such as Hurricanes [4-6].

In the past, researchers commonly used linear models for text classification because they were simple and fast. These models start by converting text into numerical representations through various encoding methods. Then, these "text embeddings" are fed into a classifier to produce the final results [2, 7]. While linear classification methods are a foundational approach widely used in text classification tasks and have achieved notable success [7], they also have some weaknesses. Despite their effectiveness, these methods face challenges related to generalization in new domains and rely on specific labels. Models like BERT, which are used as encoders in these approaches, often encounter limitations and a decline in their ability to generalize after being fine-tuned for specific classification tasks [8]. This limitation arises from the lack of utilization of label semantic information, which prevents the model from performing well in new contexts [9]. These classification methods reduce the rich semantic information present in the labels originally, turning it into basic spatial data. As a result, the models cannot fully capture all the details of the labels during training. Additionally, after training, these models can only perform specific tasks within a particular domain, and neither the number of labels nor their meanings can be changed. This limitation makes it difficult for the models to generalize well to new contexts and can even diminish the strong generalization potential of pre-trained models [8, 10].

This research is based on 'Supervised Contrastive Learning (SCL)' [11] for classifying short texts. Instead of relying on a traditional classifier, it employs a title encoder for label encoding and a separate text encoder for encoding the text itself. A contrastive learning approach is used to train both encoders, enhancing semantic comprehension and addressing domain-specific challenges. This approach effectively transforms text classification tasks into semantic understanding challenges and ensures alignment with the pre-trained model's objectives in stages of training and fine-tuning [12].

To assess the generalization capability of the supervised contrastive classification method, we use a large dataset of 1,868,703 tweets related to Hurricane Ida. Out of these, 419 tweets are manually labeled into two categories: relief

Preprint accepted at the 14th International Conference on Computer and Knowledge Engineering (ICCKE 2024). Find the published version at https://ieeexplore.ieee.org/document/10874505

evaluation and irrelevant to relief evaluation. Evaluation results show that the model reaches a classification accuracy of 0.9516 for tweets related to relief evaluation. This approach enhances the text classification models' ability to generalize and stay consistent with the task's purpose during the pre-training and fine-tuning phases, thereby preserving the model's overall ability to generalize. Additionally, the proposed model demonstrates high efficacy in supervised tasks, underscoring the considerable potential of the adopted method for practical real-world applications.

The paper is organized as follows: After this introduction, section 2 provides an overview of related research. Section 3 describes the proposed method. Section 4 includes the evaluation process, and the final section discusses the findings.

II. Related Work

In text classification, many techniques and frameworks have been created to enhance the accuracy and performance of classification models [2, 3, 13]. Powers et al. used machine learning and advanced models like BERT and XLNet to detect tweets about emergency requests during natural disasters, achieving higher accuracy than traditional methods [5]. Zhou et al. used a BERT-CNN model for detecting rescue request tweets from Hurricane Harvey, achieving an F1 score of 0.919 [6].

Recent research has investigated how incorporating label information into text classification models can improve their performance and generalization ability [9]. Zhang et al. [14] Introduced a multi-task label embedding technique that converts labels into semantic vectors, making classification a matter of vector comparison. Papas and Henderson [15] Created a versatile input-label embedding model that effectively captures complex relationships between labels. Their method demonstrates strong performance with both familiar and novel labels. Kuoang and Davison [16] proposed class-specific word embeddings using linear combinations to incorporate class information, enhancing topic and sentiment classification accuracy. These methods leverage label meanings and global information, improving traditional linear classification models and creating more effective text classification systems.

Recent studies have focused on using contrastive learning for crisis tweet classification to improve generalization and performance. In this regard, Nguyen and Rudra introduced an interpretable classifier that sorts tweets into humanitarian categories and extracts explanatory segments. These explanatory segments provide additional information and necessary explanations that help better understand the tweet’s categorization and analysis [17]. Alongside, Xie et al. developed a supervised contrastive learning framework designed for handling multi-label text classification in essential situations [18].

Ghosh et al. introduced a graph-based method for supervised contrastive learning, which leverages labeled data from related tasks, and has shown improved effectiveness across different domains [19]. These studies demonstrate that contrastive learning can enhance crisis tweet classification by addressing multi-label, cross-domain, and few-shot learning challenges. This technique, gaining attention in NLP and computer vision, improves embeddings by pulling similar feature vectors together and pushing dissimilar ones apart [11]. Yi Liang et al. introduced a semantic-based comparative classification method using two encoders for labels and text. Their approach focuses on sample similarities and differences, improving model generalization and accuracy. Trained on diverse datasets like news headlines, question answering, and text matching, it outperforms traditional models and adapts better to new data [8].

The contribution of this paper is the use of supervised contrastive learning to classify tweets about relief efforts during crises like Hurricane Ida. Unlike traditional methods which focus on encoding textual features, this approach utilizes contrastive learning to enhance accuracy and generalization.

III. Proposed method

Supervised Contrastive Learning (SCL) is a cutting-edge technique designed to enhance text representations and improve classification performance. By leveraging label information, this method helps distinguish between samples and emphasizes unique features across different categories [11]. Unlike conventional approaches that depend on prototypes to assess similarity, Supervised contrastive learning refines text representations through direct use of labels. The framework is built around two key elements: the "Support Set," which supplies examples to facilitate the learning process, and the "Query Set," which consists of instances that the model should classify. The purpose of the model is to bring the text representations with identical labels closer together while separating those with distinct labels [20]. This results in stronger semantic representations that better differentiate between categories. Figure 1 presents an overview of the proposed method.

In this framework, contrastive learning, traditionally employed in self-supervised scenarios, is adapted for a fully supervised setting to enhance the efficiency of using labeled data. This method can significantly improve the accuracy and discriminative power of the model in text classification tasks. Contrastive learning, in general, is applicable in any context where distinctions or contrasts exist. It involves comparing the similarity between texts and their labels.

This paper introduces a classification method grounded in contrastive learning that employs semantic vectors of texts and labels. Suitable encoders are utilized to encode both texts and labels. Based on semantic similarity, the correct label for each text is predicted. This study focuses on classifying tweets from the Hurricane Ida dataset, which is divided into two categories: "Tweets related to relief evaluation" and "Tweets unrelated to relief evaluation." Initially, 419 tweets related and unrelated to relief evaluation were collected. Our approach to categorizing these tweets involves several key steps:

A. Data preprocessing

In the initial stage, the raw tweet data is thoroughly cleaned and normalized by removing textual noise—such as emojis, URLs,

Preprint accepted at the 14th International Conference on Computer and Knowledge Engineering (ICCKE 2024). Find the published version at https://ieeexplore.ieee.org/document/10874505

and unnecessary symbols—to reduce distractions and inconsistencies. The text is then standardized by converting it to lowercase and ensuring uniform punctuation, optimizing the encoding process.

B. Encoding with BERT

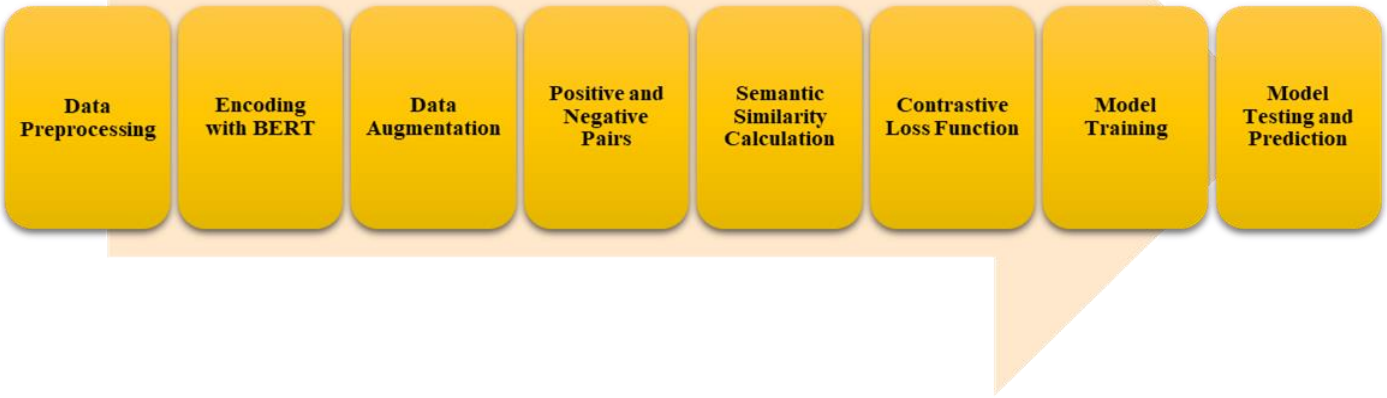

Fig. 1. Supervised Contrastive Learning (SCL) framework for text classification

The title encoder processes labels or titles associated with tweets by employing a pre-trained BERT model. Similarly, the text encoder processes the tweet content independently, with BERT extracting semantic and syntactic features from the text and converting them into numerical vectors. This system enables the model to gain a deeper understanding of the structure and meaning of the text.

C. Data augmentation

This stage aims to increase data diversity to enhance the generalization and performance of the model. This can be particularly effective when the number of labeled samples is limited. Techniques such as word shuffling, random removal of certain vocabulary, or replacing words with synonyms are employed to strengthen the data and prevent overfitting. For instance, an augmented version of "The help arrived very

timely, and we are satisfied with their assistance." might be: "The help arrived very quickly and we are satisfied with their assistance." This stage involves synonym replacement, random word removal, and morphological alterations, which help increase the diversity of the training data. It also enables the model to generalize more effectively. Positive pairs consist of the original tweets and those generated from the data augmentation process, while negative pairs include tweets that are unrelated to relief evaluation.

D. Formation of positive and negative pairs

The model utilizes positive and negative pairs to enhance learning: positive pairs consist of tweets that belong to the same category. For instance, the tweet "Thank you, the rescuers arrived quickly" and its augmented version (e.g., "Thank you, the rescuers came quickly") form a positive pair. Negative pairs are comprised of tweets that belong to different categories, such as the tweets "Thanks to the medical relief team that arrived quickly" (related to relief evaluation) and "It rained yesterday" (unrelated to relief evaluation). The model must learn that positive pairs exhibit high semantic similarity, while negative pairs show lower similarity.

E. Semantic similarity calculation

In this phase, the feature vectors and labels for each tweet, obtained from BERT, are used to find semantic similarity between them. Semantic similarity is assessed using cosine similarity. This metric indicates how closely two vectors align in the vector space. If two positive tweets have a high cosine similarity (close to 1), it means they have similar meanings. On the other hand, for negative pairs, the model tries to get a low cosine similarity (around 0).

F. Contrastive loss function

The model utilizes a contrastive loss function to reduce the cosine distance between positive pairs and increase it between negative pairs. This mechanism has allowed the model to enhance its capability to differentiate relevant tweets from unrelated ones over time. The loss function encourages the model to increase similarity for positive pairs and decrease it for negative pairs. The goal is to bring similar tweets closer together and move unrelated tweets further apart. The model is trained using positive and negative pairs. We used the AdamW algorithm for optimization with a learning rate of 0.00001. The training batch size is 16, and the test batch size is 64. The model was trained for 20 epochs.

G. Model testing

After completing the training phase, the model has been tested using new tweets. For each new tweet, the model calculates its cosine similarity with various labels (either relevant or irrelevant) and selects the label with the highest similarity as the final classification. For example:

Preprint accepted at the 14th International Conference on Computer and Knowledge Engineering (ICCKE 2024). Find the published version at https://ieeexplore.ieee.org/document/10874505

If the new tweet, "The relief services were outstanding; we are satisfied," is given to the model, it recognizes it as a tweet about relief evaluation.

H. Benefits and model performance

This methodology enhances the model's generalization by emphasizing semantic comprehension, allowing effective adaptation to various contexts, even with new data. The integration of contrastive learning boosts reliability across diverse tasks. Using BERT to encode titles and texts, the approach efficiently classifies tweets on their relevance to relief evaluations during Hurricane Ida, achieving better accuracy. Contrastive learning significantly aids in differentiating between tweets related to relief assessments and those that are not by analyzing positive and negative pairs.

Contrastive learning encourages the model to group similar tweets while separating different ones through the contrastive loss function, assigning high similarity to relief-related tweets and low similarity to unrelated ones. For example, "The rescuers arrived promptly" and "The saviors came quickly" are positive pairs, while "We're dissatisfied with the aid" and "It was rainy today" are negative pairs. In summary, contrastive learning enhances tweet classification accuracy by reinforcing the distinctions between related and unrelated tweets, which leads to improved model performance.

IV. Evaluation

To evaluate the proposed model, we used a dataset of tweets about Hurricane Ida. This dataset has 1,868,703 tweets, including 419 tweets that were manually labeled. We divided the tweets into two groups: "tweets about relief evaluation" and "tweets not about relief evaluation". The model uses supervised contrastive learning to classify the tweets. We assess its performance using several metrics.

Accuracy: This measure shows how the model's predictions are accurate by comparing the number of correctly identified tweets to the total number of tweets. The model achieved an accuracy rate of 0.9516, meaning 95.16% of the tweets were classified correctly.

Precision: This metric measures the proportion of accurately identified positive cases out of all the positive labels given by the model. With a precision of 0.9622, the model successfully recognized 96.22% of the tweets, marked as relevant to the relief evaluation.

Recall: This metric measures the percentage of actual positive cases that the model identified correctly. With a recall score of 0.9272, that means the model recognized 92.72% of the relevant tweets.

F1-Score: This metric calculates the harmonic average of precision and recall to provide a balanced view of the model's effectiveness. An F1-Score of 0.9444 shows that the model performs very well in both precision and recall.

Table 1 compares the results of two classification models: SVM and contrastive learning. The contrastive learning model outperforms SVM in terms of accuracy, achieving 0.9516 compared to 0.92, indicating better overall performance in predicting the correct classification of tweets. Both models have similar recall scores, but contrastive learning is a bit better than SVM (0.9272 vs. 0.92), showing a small improvement in finding relevant tweets. However, the contrastive learning model has a much higher precision score of 0.9622, while SVM's is 0.92, meaning it makes fewer mistakes by classifying irrelevant tweets as relevant. The F1-score, which balances precision and recall, is higher in the contrastive learning model (0.9444 vs. 0.92), showing that it has better overall performance. Overall, the contrastive learning model outperforms SVM in all key metrics, especially in precision and F1-score, proving its effectiveness in text classification tasks. Table 1 compares the results.

Table 1. Comparison between the proposed method and SVM

Models	Accuracy	Recall	Precision	F1-score
SVM	0.92	0.92	0.92	0.92
Contrastive learning	**0.9516**	0.9272	0.9622	0.9444

V. Conclusion

The main advancement of this work is the use of contrastive learning techniques to encode both labels and texts, which has not been studied in this context before. We applied this model specifically to classify tweets about Hurricane Ida in order to identify tweets relevant to relief evaluation from a large dataset, where only a small percentage of tweets were related to relief efforts. This task is important for helping organizations such as the Global Red Cross, the United Nations, and other relevant agencies make better decisions and improve aid delivery during crises and natural disasters.

The proposed approach has demonstrated strong performance, achieving high accuracy and precision scores along with substantial recall and F1 scores, in detecting tweets related to relief evaluation. This effectiveness demonstrates its capability to handle new and varied data while overcoming limitations associated with traditional methods, such as the dependency on exact labels and challenges in generalizing to new contexts. The approach also utilizes contrastive learning to tackle situations where only a small amount of labeled data is available; in our case, we only had 419 labeled tweets.

Additionally, this study underscores the considerable promise of supervised contrastive learning in advancing the accuracy and robustness of text classification models. Future investigations might involve applying this technique to different natural language processing tasks and examining its performance with diverse types of textual data and label variations.

Acknowledgment

We wish to convey our sincere appreciation to the ICT Research Institute for their essential support during this research.

References

Preprint accepted at the 14th International Conference on Computer and Knowledge Engineering (ICCKE 2024). Find the published version at https://ieeexplore.ieee.org/document/10874505

[1] D.W. Otter, J.R. Medina, and J.K. Kalita, "A survey of the usages of deep learning for natural language processing", IEEE Trans. Neural Networks Learn. Syst., vol. 32, no. 2, pp. 604-624, 2020.

[2] M.M. Mirończuk and J. Protasiewicz, "A recent overview of the state-of-the-art elements of text classification", Expert Syst. Appl., vol. 106, pp. 36-54, 2018.

[3] D. Rogers, A. Chen, and J. Smith, "Real-time text classification of user-generated content on social media: Systematic review", IEEE Trans. Comput. Soc. Syst., vol. 9, no. 4, pp. 1154-1166, 2021.

[4] L. Zou, A. Gupta, and J. Lee, "Social media for emergency rescue: An analysis of rescue requests on Twitter during Hurricane Harvey", Int. J. Disaster Risk Reduction, vol. 85, p. 103513, 2023.

[5] C.J. Powers, A. Carter, and L. Johnson, "Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach", Int. J. Inf. Manag. Data Insights, vol. 3, no. 1, p. 100164, 2023.

[6] B. Zhou, C. Yang, and J. Wang, "VictimFinder: Harvesting rescue requests in disaster response from social media with BERT", Comput. Environ. Urban Syst., vol. 95, p. 101824, 2022.

[7] Y.-C. Lin, J. Doe, and M. Lee, "Linear classifier: An often-forgotten baseline for text classification", arXiv preprint arXiv:2306.07111, 2023.

[8] Y. Liang, T. Tohti, and A. Hamdulla, "Contrastive classification: A label-independent generalization model for text classification", Expert Syst. Appl., vol. 245, p. 123130, 2024.

[9] X. Liu, Y. Zhang, and S. Chen, "Label-guided learning for text classification", arXiv preprint arXiv:2002.10772, 2020.

[10] A. Mueller, T. Wang, and B. Smith, "Label semantic aware pre-training for few-shot text classification", arXiv preprint arXiv:2204.07128, 2022.

[11] P. Khosla, S. Gupta, and R. Choudhary, "Supervised contrastive learning", Adv. Neural Inf. Process. Syst., vol. 33, pp. 18661-18673, 2020.

[12] A. Mahabal, R. Jain, and S. Patel, "Text classification with few examples using controlled generalization", arXiv preprint arXiv:2005.08469, 2020.

[13] Q. Li, J. Zhao, and X. Liu, "A survey on text classification: From shallow to deep learning", arXiv preprint arXiv:2008.00364, 2020.

[14] H. Zhang, S. Lee, and Y. Chen, "Multi-task label embedding for text classification", arXiv preprint arXiv:1710.07210, 2017.

[15] N. Pappas and J. Henderson, "Gile: A generalized input-label embedding for text classification", Trans. Assoc. Comput. Linguist., vol. 7, pp. 139-155, 2019.

[16] S. Kuang and B.D. Davison, "Class-specific word embedding through linear compositionality", in 2018 IEEE Int. Conf. Big Data Smart Comput. (BigComp), 2018, pp. 123-126.

[17] T.H. Nguyen and K. Rudra, "Rationale aware contrastive learning based approach to classify and summarize crisis-related microblogs", in Proc. 31st ACM Int. Conf. Inf. Knowl. Manag., 2022.

[18] S. Xie, T. Chen, and Y. Liu, "Multi-label disaster text classification via supervised contrastive learning for social media data", Comput. Electr. Eng., vol. 104, p. 108401, 2022.

[19] S. Ghosh, S. Maji, and M.S. Desarkar, "Effective utilization of labeled data from related tasks using graph contrastive pretraining: Application to disaster related text classification", in Proc. 37th ACM/SIGAPP Symp. Appl. Comput., 2022.

[20] J. Chen, K. Zhang, and P. Singh, "ContrastNet: A contrastive learning framework for few-shot text classification", in Proc. AAAI Conf. Artif. Intell., 2022.

[21] J. Kim, Y. Park, and H. Lee, "Generalized supervised contrastive learning", arXiv preprint arXiv:2206.00384, 2022.

[22] Y. Zhang, S. Wu, and L. Chen, "Metadata-induced contrastive learning for zero-shot multi-label text classification", in Proc. ACM Web Conf., 2022.