A BERT-based with fuzzy logic sentimental classifier for sarcasm detection

Tianyou Wang

Faculty of Science and Technology, BNU-HKBU United International College, Zhuhai, China

wtywpzyt@163.com

Abstract. This study explores the challenging task of sarcasm detection in text, a crucial aspect of sentiment analysis in Natural Language Processing (NLP). Sarcasm, characterized by irony and nuanced language, often complicates the interpretation of emotional tone in text. To address this, the study employs a hybrid model that integrates BERT (Bidirectional Encoder Representations from Transformers) with fuzzy logic. BERT's deep semantic understanding is combined with the sequential processing of LSTM and the flexible decision-making capabilities of fuzzy logic. The model's performance is validated using multiple datasets, demonstrating a significant improvement in accuracy, particularly in detecting subtle and context-dependent sarcasm. This research contributes to the advancement of sentiment analysis, offering a robust framework for handling complex linguistic expressions in various NLP applications.

Keywords: Sarcasm Detection, Transformer, BERT, Fuzzy Logic, Sentiment Analysis.

1. Introduction

In today's era of information overload, people frequently express opinions, emotions, and ideas through text. However, detecting the true emotional tone in text, particularly in complex expressions like sarcasm, presents a significant challenge. Sarcasm, a form of rhetorically clever language, conveys irony or playfulness through words that often obscure their true meaning. For a machine learning model to accurately detect sarcasm, it must grasp underlying semantics through extensive training, which is a key difficulty in sentiment analysis. Additionally, the model must be highly sensitive to context, nuance, and metaphor.

The application of sarcasm detection in text has broad implications. In social media sentiment analysis, it can assist companies in better understanding the authentic emotional responses of users. In news sentiment analysis, it can identify irony or exaggeration, thereby improving the accuracy of information interpretation. Furthermore, sarcasm detection can enhance the sentiment analysis of product reviews, providing companies with a more nuanced understanding of user feedback.

The Transformer model, with its attention mechanism, forms the foundation of BERT (Bidirectional Encoder Representations from Transformers) [1]. This model assigns varying importance to different parts of the input, focusing on critical aspects by using self-attention to capture long-range dependencies between words. Multi-head self-attention further refines this by processing information across multiple semantic dimensions. BERT advances the Transformer model by incorporating a bidirectional encoder, which allows for a deeper understanding of context. BERT processes input through token, segment, and position embeddings, transforming entire sentences into rich, high-dimensional vectors. When combined with fuzzy logic, which manages uncertainty and partial truths, these models enable more nuanced decision-making and inference in sentiment analysis [2].

The study aims to analyze sarcasm in text using fuzzy logic, determining whether a given text contains sarcastic sentiment. Unlike classical logic systems that rely on binary decisions, fuzzy logic models imprecise reasoning patterns, mirroring the human ability to make rational decisions in uncertain environments. Integrating fuzzy logic thus facilitates a more comprehensive and flexible approach to sentiment analysis in text.

2. Methods

2.1. Bert Model

BERT (Bidirectional Encoder Representations from Transformers) is a prominent model in the field of Natural Language Processing (NLP), typically trained using large-scale text corpora that are not specific to any particular NLP task. The goal of such pre-training is to make the model's output adequately representative of textual semantics, enabling it to perform well across a range of downstream tasks.

The pre-training process involves adjusting the model's parameters to accurately capture the semantics of the text, which is crucial for enhancing the performance on subsequent NLP tasks. BERT employs two unsupervised learning strategies for this purpose: the masked language model and next sentence prediction [3].

The masked language model (MLM) combines the attention mechanism with contextual information to enable BERT's bidirectional feature representation. In this approach, the model randomly masks 15% of the words in a sentence using a special tag, "[MASK]." The model then predicts the masked words based on the surrounding context. However, since actual downstream tasks do not include masked words, BERT employs a strategy to ensure consistency. Specifically, 80% of the time, the "[MASK]" tag is replaced by the original word, 10% of the time it is replaced by a random word, and 10% of the time, no replacement occurs. This strategy forces the model to rely on full contextual information when predicting words, thereby enhancing its understanding of semantics.

The next sentence prediction (NSP) task further aids in making the model adaptable to downstream fine-tuning. During training, BERT is provided with pairs of sentences: 50% of the time, the sentences are sequential and related, while the other 50% of the time, the second sentence is randomly selected and unrelated to the first. The model then determines whether the two sentences are related or not. This task helps BERT develop a deeper understanding of the relationships between sentences, which is critical for tasks such as question answering and natural language inference.

2.2. Fuzzy Logic

In fuzzy logic, the primary task during inference is to evaluate the outcomes generated by each rule based on the provided fuzzy input and then combine these outcomes to reach a final decision. When it comes to defuzzification, the goal is to convert the fuzzy output into a precise numerical value that can be used as the final output.

Two common methods for performing inference in fuzzy logic systems are Mamdani fuzzy inference and Sugeno fuzzy inference. In Mamdani fuzzy inference, the output is still a fuzzy set, which maintains the ambiguity inherent in the input data. This method is widely used due to its intuitive approach and the ability to model systems where the output can be expressed in linguistic terms.

On the other hand, Sugeno fuzzy inference produces an output that is a constant or a linear function, defined by a singleton membership function. The singleton membership function is unique in that it has a non-zero value at only one specific point in its domain, while the value is zero elsewhere. This characteristic makes Sugeno inference particularly useful in control systems and optimization problems, where a crisp, numerical output is required for practical implementation.

2.3. Long Short-Term Memory (LSTM)

RNNs treat input sequences uniformly, where each unit's output influences the next unit's input. However, LSTMs and GRUs improve on RNNs by managing long-term dependencies more effectively. LSTMs, a type of RNN, address the challenge RNNs face with long sequences by using mechanisms like the Forget Gate. This gate selectively discards or retains information from the cell state, helping manage memory. It applies a sigmoid function to decide the extent of retention, with 0 meaning complete discard and 1 meaning full retention.

The Input Gate in LSTMs decides what new information to store, using a candidate cell state to determine which memories are useful. It adds valuable new information to the cell state, integrating both retained old memories and new memories. Finally, the Output Gate determines the final output, which is the new cell state.

2.4. Implementation Procedure

The approach employs a hybrid integrated method that leverages a combination of BERT, LSTM, and fuzzy logic to provide a highly specific solution for sarcasm detection (Figure 1). To determine whether a text is sarcastic, BERT is first used to extract deep semantic features from the text. These features are then processed by an LSTM network, which captures the sequential dependencies and contextual nuances within the text [4].

Finally, a fuzzy logic layer reclassifies the output from BERT and LSTM according to predefined fuzzy rules. This additional fuzzy logic classification enhances the model's ability to accurately identify sarcastic and ambiguous statements, particularly those that might be missed by traditional models [5]. The Sarcasm Corpus V2 Dataset [6], SARC Dataset [7], Headlines Dataset [8] are utilized for this study to train and evaluate the model's performance.

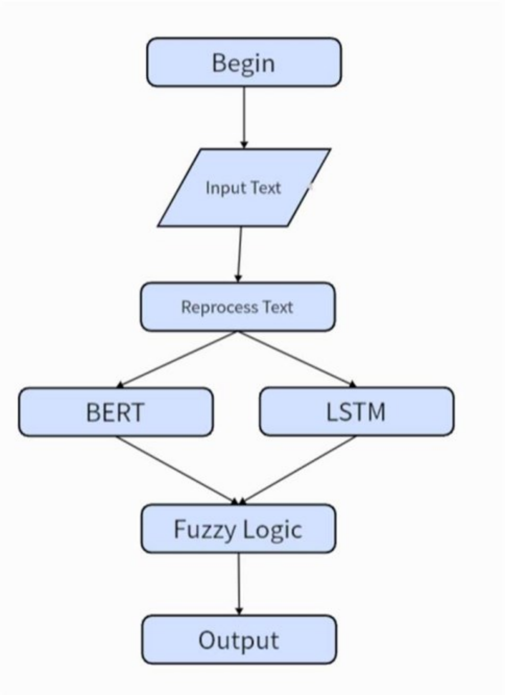

Figure 1. Flowchart of proposed framework

2.5. Dataset Processing

The Internet Argument Corpus was selected, focusing on the collection and analysis of online disputes and conflicts. This corpus captures opinions, statements, and arguments on various topics, highlighting interactions and debates among participants. It categorizes sentences into three types: general statements (GEN), exaggerations (HYP), and rhetorical questions (RQ). A selection of 9,386 sentences was made, consisting of 6,520 GEN, 1,164 HYP, and 1,702 RQ, including 4,800 satirical and 4,586 non-satirical sentences. Each sentence was labeled as ironic (1) or non-ironic (0). Using Cross-Validation, the data was split into 6,570 training and 2,816 test set instances.

2.6. BERT and Fuzzy Logic Settings

BERT: The pretrained 'Bert-base-uncased' model is used, with a single-layer feedforward classifier comprising two fully connected layers, connected to an external fully connected layer. Optimization is handled by the SGD optimizer with a learning rate of 1e-4, and CrossEntropyLoss() is employed as the loss function. Statements are tokenized and trained using the training set over ten epochs, with parameters saved for future use. During prediction, these trained parameters are applied to the test set, and predicted labels are compared with actual outcomes to optimize accuracy and minimize errors.

Fuzzy Logic: The fuzzy logic module uses outputs from BERT and LSTM to blur the test set text, performing fuzzy inference based on predefined rules. The defuzzifier produces a probability indicating whether a statement is ironic, with high probabilities labeled as ironic (1) and low as non-ironic (0). This module enhances precision. Fuzzy logic is applied to analyze ambiguous language, supported by a fuzzy control system for reliable decision-making. The fuzzifier converts input values into membership degrees for different classes.

The fuzzy rules are shown as below:

- IF (LSTM predicts False) AND (BERT predicts True), THEN (Degree of Membership is low).
- IF (LSTM predicts False) AND (BERT predicts True) OR (LSTM predicts True) AND (BERT predicts False), THEN (Degree of Membership is medium).
- IF (LSTM predicts True) AND (BERT predicts True), THEN (Degree of Membership is high).

Logical calculations use the minimum membership method (Conjunction) and maximum membership method (Disjunction). Fuzzy inference is performed based on these rules, and the results are de-fuzzified using the centroid method to determine whether a statement is ironic, thereby improving accuracy.

3. Results

The model's accuracy was evaluated before and after the integration of fuzzy logic. We compared the results and plotted the curves for loss, training accuracy, and test accuracy over 30 iterations, as depicted in Figures 2 and 3. Initially, the model achieved an accuracy of 76.9%. After incorporating fuzzy logic, the accuracy increased to 82.2%, representing a 5.3% improvement.

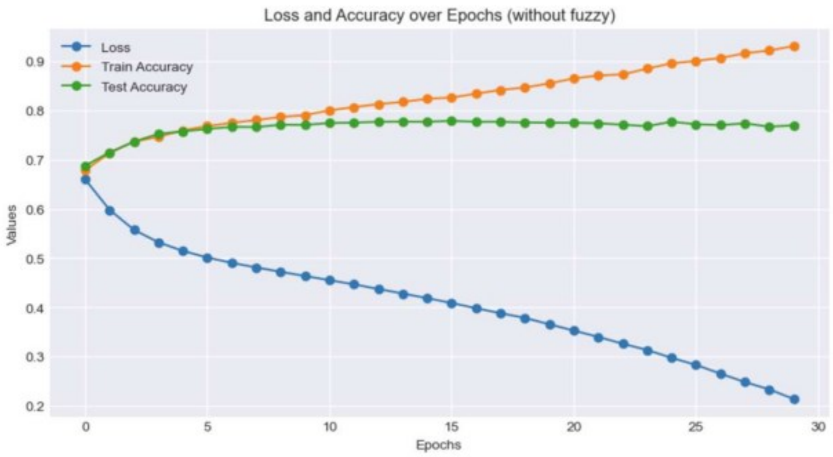

Figure 2. Loss and Accuracy over Epochs (without fuzzy)

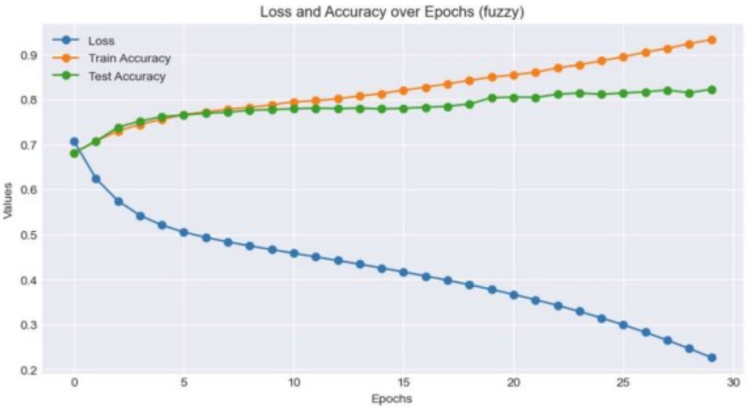

Figure 3. Loss and Accuracy over Epochs (fuzzy)

To assess the performance of the model, the most common classification metrics were calculated: accuracy, precision, recall, and F1 score. Additionally, the confusion matrices for the evaluations are displayed in Figures 4 to 6. The model's performance was compared with other existing models, using three different datasets—Sarcasm Corpus V2, SARC, and Headlines—to enhance its generalization ability and robustness.

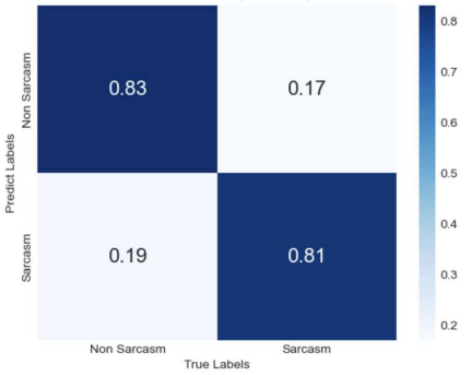

Figure 4. Confusion matrix obtained using the Sarcasm Corpus V2 dataset.

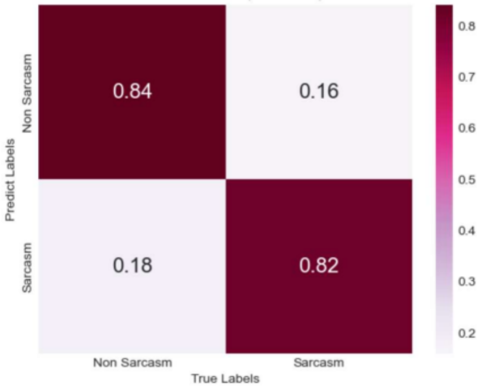

Figure 5. Confusion matrix obtained using the SARC dataset

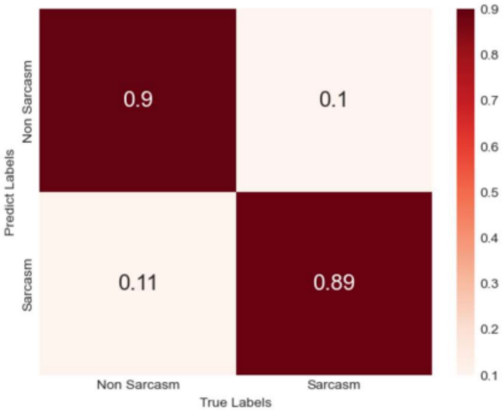

Figure 6. Confusion matrix obtained using the Headline dataset

4. Discussion

This study has some limitations. The datasets used primarily consist of short texts within the comment category, leaving the model's performance on longer textual content unexplored. Addressing this limitation will be a key direction for future research. Despite this, the study's significance lies in its innovative approach to sarcasm detection, which improves the accuracy of deep learning models in natural language tasks. The proposed research framework contributes positively to the development of sentiment analysis, the enhancement of sentiment intelligence processing, and the expansion of research methodologies in other NLP domains [9].

5. Conclusion

This study addresses the complex task of sarcasm detection within the field of NLP by employing an approach that integrates BERT modeling with fuzzy logic techniques. The research aims to enhance the accuracy and adaptability of sarcasm detection models, with a focus on the subtle and context-dependent nature of sarcastic expressions. A comparative analysis with existing methods underscores the strengths of this approach. The model demonstrates robust performance across a variety of datasets, effectively handling the complexities and nuances of sarcasm in different contexts. This consistency highlights the model's applicability and effectiveness in sarcasm detection.

References

[1] Sharma, D. K., Singh, B., Agarwal, S., Pachauri, N., Alhussan, A. A., & Abdallah, H. A. (2023). Sarcasm Detection over Social Media Platforms Using Hybrid Ensemble Model with Fuzzy Logic. Electronics, 12(4), 937. https://doi.org/10.3390/electronics12040937

[2] Zadeh, L. A. (1988). Fuzzy logic. Computer, 21(4), 83-93. https://10.1109/2.53

[3] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. https://arxiv.org/abs/1810.04805

[4] Cai, R., Qin, B., Chen, Y., Zhang, L., Yang, R., Chen, S., & Wang, W. (2020). Sentiment Analysis About Investors and Consumers in Energy Market Based on BERT-BiLSTM. IEEE Access, 8, 171408-171415. https://doi.org/10.1109/ACCESS.2020.3024750

[5] Ansari, G., Shukla, S., Gupta, M., & Gupta, H. (2022). A Fuzzy Approach for Opinion Summarization of Product Reviews. In 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), 1-6.

[6] Shereen Oraby, Vrindavan Harrison, Lena Reed, Ernesto Hernandez, Ellen Riloff, and Marilyn Walker. (2016). Creating and Characterizing a Diverse Corpus of Sarcasm in Dialogue. In The 17th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL), Los Angeles, California, USA.

[7] Khodak, M., Saunshi, N., & Vodrahalli, K. (2017). A Large Self-Annotated Corpus for Sarcasm. arXiv preprint arXiv:1704.05579. https://arxiv.org/abs/1704.05579

[8] Misra, R., & Arora, P. (2019). Sarcasm Detection using Hybrid Neural Network. arXiv preprint arXiv:1908.07414. https://arxiv.org/abs/1908.07414

[9] Potamias, R. A., Siolas, G., & Stafylopatis, A. (2020). A transformer-based approach to irony and sarcasm detection. Neural Computing & Applications, 32(12), 17309–17320. https://doi.org/10.1007/s00521-020-05102-3