Research on Sarcastic Emotion Recognition Based on Multiple Feature Fusion

Kaihao Si

Cyberspace Security Major, Northwestern Polytechnical University, 710129 Xi'an, China

Abstract. Sarcasm detection significantly enhances the performance of various natural language processing applications, such as sentiment analysis, opinion mining, and stance detection. Despite considerable advancements in this field, research results remain fragmented across diverse datasets and studies. This paper offers a critical review of two predominant models in sarcasm detection. The first model utilizes BERT within an intermediate task transfer learning framework, leveraging the connection between sarcasm and underlying negative emotions and sentiments. This model enhances the sarcasm detection capability through a strategic knowledge infusion into the transfer learning process. The second model reviewed deploys a multi-head attention-based bidirectional LSTM architecture. This approach incorporates pre-trained word embeddings, multi-head attention mechanisms, and custom-crafted features to proficiently identify sarcasm in social media datasets. Comparative assessments on standard datasets reveal that both models achieve superior performance over many existing approaches in the field. At last, this paper explores the direction for future improvement based on the conclusions.

1 Introduction

Accurate and fast processing of natural language sentiment analysis is a hot topic in current research, and one of the important challenges is the detection of sarcastic language. Sarcasm is a complex linguistic phenomenon, and its actual meaning is often the opposite of its literal meaning. In daily life, sarcasm is also a common way of language expression. People usually use sarcasm to euphemistically express unpleasant emotions or make jokes. A study based on Reddit showed that approximately 23% of comments on the platform were marked as sarcastic [1]. These sarcastic comments mainly appear in discussions about politics, social hot issues or other controversial topics. Generally speaking, the forms of satirical expression are mainly reflected in the following three aspects: (a) contrast in part of speech; (b) contrast in context; and (c) contrast in structure. Therefore, there is research literature that uses sentiment information in sentences to detect sarcasm based on feature engineering [2,3]. To further explain these three aspects, this article gives an example:

Exp 1. I love when all your teachers throw assignments and projects at you on dead week and they still expect you to be studying for their final 🙄.

Corresponding author: 13010220040115181 6@mail.nwpu.edu.cn

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

lexical contrast: The word "love" is used as an irony here, expressing a feeling that is opposite to its literal meaning. The " 🙄 " emoticon at the end of the sentence is an expression of rolling eyes, which is usually used to express helplessness, irritability or sarcasm, further strengthening the emotional expression of sarcasm.

Contextual contrast: Students are expected to focus on revision before an exam, but are distracted by a large number of assignments and projects, which is the opposite of what is normally expected.

structural contrast: Although there is no dramatic irony involved in this sentence, if the reader knows from other background information that the teachers don't actually care about the students' burdens, then this can also create a certain degree of structural contrast.

From this example, these three aspects can cover sarcastic language in most cases, and through this aspect, the algorithms effectively detect sarcasm in text data.

Current algorithms for sarcastic language detection fall into two main categories: feature engineering methods and deep learning approaches. Feature engineering techniques often involve the manual construction of features for sarcasm identification. This process, while detailed, is time-intensive and does not generalize well across different contexts. In contrast, deep learning strategies enable the automatic extraction of relevant features, enhancing efficiency and adaptability. In the realm of feature engineering, Bharti et al. introduced two innovative algorithms for sarcasm detection [3]. The first is based on parsing to generate a dictionary specifically tailored for sarcasm, while the second focuses on recognizing sarcasm through interjections. Additionally, Farías et al. employed semantic features derived from the similarity and sentiment linkage found in semantic resources like AFINN, SentiWordNet, and General Inquirer [4]. Transitioning to deep learning methods, Ghosh and Veale developed a composite neural network that integrates CNNs and LSTMs for enhanced sarcasm detection [5]. They further utilized LIWC to analyze author sentiment, blending semantic and contextual data [6]. Their approach features a two-layer network combining CNNs with dual LSTM models, grounded in word embeddings. Highlighting the importance of label sentiment in detecting sarcasm, Maynard et al. presented key insights into its crucial role [7]. Additionally, Ren et al. introduced two advanced context-enhanced neural models crafted specifically for this purpose [8].

Recently, sarcasm detection has also been explored through multimodal approaches. Cai et al. merged text, image, and attribute modalities within a hierarchical fusion model tailored for sarcasm identification [9]. Furthermore, Li et al. demonstrated the efficacy of integrating textual and visual data, proving that such multimodal methods adeptly handle the complexity and subtlety inherent in sarcasm [10]. This array of methodologies underlines the diverse strategies researchers have deployed to refine sarcasm detection, catering to its intricate nature and the variety of its manifestations. Wang et al. proposed a method to bridge text and image information, using independent text and image models to handle the sarcasm detection problem without the need for pre-training on text and image paired data [11]. They independently extracted text and image features and then combined the two through a "bridge" mechanism to achieve sarcasm detection. To address the lack of datasets in sarcasm detection, Zhang et al. proposed an adversarial learning framework to improve robustness and generalization ability [12].

While significant advancements have been made in the field of sarcasm detection, research results remain fragmented across various datasets and different investigations. In contrast to previous works, this current study focuses predominantly on evaluating two distinct models of sarcasm detection. Firstly, it examines the integration of BERT for sarcasm detection through intermediate task transfer learning. This model leverages the intrinsic relationship between sarcasm and subtler negative emotional cues. It innovatively applies sentiment classification and detection as standalone intermediary tasks, thereby enriching the primary task of sarcasm detection with nuanced affective information. Secondly, the study

delves into a model employing a bidirectional LSTM enhanced by multi-head attention mechanisms. Utilizing pre-trained word embeddings and meticulously crafted features, this model proves to be highly effective in recognizing sarcasm within social media contexts.

2 Models overview

This paper reviews two sarcasm detection models suitable for different situations to more effectively identify sarcastic texts. The two models respectively demonstrate the current state-of-the-art BURT pre-trained language model based on transfer learning and the bidirectional LSTM model based on multi-head attention mechanism, aiming to comprehensively improve the accuracy and robustness of sarcasm detection [13,14]. The design concepts and implementation methods of the two models are described in detail below.

2.1 BURT pre-trained language model based on transfer learning

2.1.1 BERT pre-trained language model

The BERT language model, a pre-trained tool, has demonstrated substantial improvements across various tasks related to understanding natural language. This model was specifically tailored for identifying sarcasm by employing the bert-base-uncased version from the Hugging Face Transformers library, which underwent fine-tuning processes. Additionally, a linear classifier was integrated with the BERT framework, leveraging the ultimate hidden state of the [CLS] token to facilitate sentence classification.

2.1.2 Intermediate Task Transfer Learning

This study proposes a transfer learning framework involving emotion classification and sentiment detection as intermediate tasks to assess their impact on enhancing the BERT model's efficacy in sarcasm detection. The methodology is illustrated in Fig. 1, outlining the sequential steps in implementing the framework.

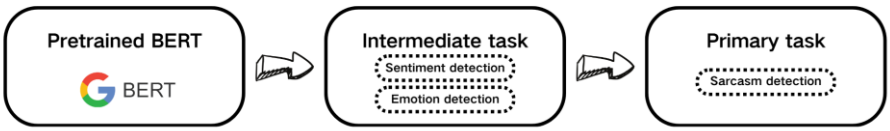

Fig. 1. Transfer Learning Framework(Photo/Picture credit : Original)

For the task of sarcasm detection, the objective is to discern whether a piece of text, such as a tweet, message, comment, or sentence, contains sarcasm based solely on the textual information provided.

Regarding sentiment detection, the EmoNet dataset, which compiles Twitter posts annotated through distant supervision with the Plutchik-24 sentiment set, is utilized. This dataset specifically includes a vocabulary that is rich in emotions, with tweets categorized according to the Plutchik-8 sentiment model, which encompasses joy, surprise, trust, anticipation, sadness, fear, anger, and disgust. For the purposes of this analysis, emotions linked to negative connotations (sadness, fear, anger, and disgust) are grouped under the negative sentiment category (0), while emotions such as joy, surprise, trust, and anticipation are assigned to positive categories. In this study, the BERT model was initially trained on the

EmoNet dataset in a supervised format. Subsequently, this enhanced model was employed in the specific investigation of sarcasm detection, by further fine-tuning on related textual data. This two-step training approach ensures the model effectively adapts to the nuances involved in recognizing sarcastic content.

2.2 Bidirectional LSTM model based on multi-head attention

In Reference, the authors developed a sophisticated neural network architecture combining multi-head self-attention with a dual-direction long short-term memory (BiLSTM) system [14]. Unlike traditional recurrent neural networks (RNNs), LSTMs are adept at preserving long-term data dependencies, efficiently bypassing issues related to gradient decay or explosion, as outlined in Reference [15]. The BiLSTM framework integrates a forward LSTM layer, which processes historical sequence data, and a backward LSTM layer for future sequence insights. Both layers converge on a common output layer, enhancing the network's predictive accuracy. The model further incorporates a multi-head attention mechanism within the bidirectional LSTM structure, enabling simultaneous consideration of various information sources from distinct representational areas at differing sequence positions. This approach results in the multi-head attention-based BiLSTM (MHA-BiLSTM), which is organized into five principal components: the word embedding layer, character encoding layer, sentence-level multi-head attention layer, an auxiliary feature connection, and a SOFTMAX layer, as depicted in Fig. 2 of their work.

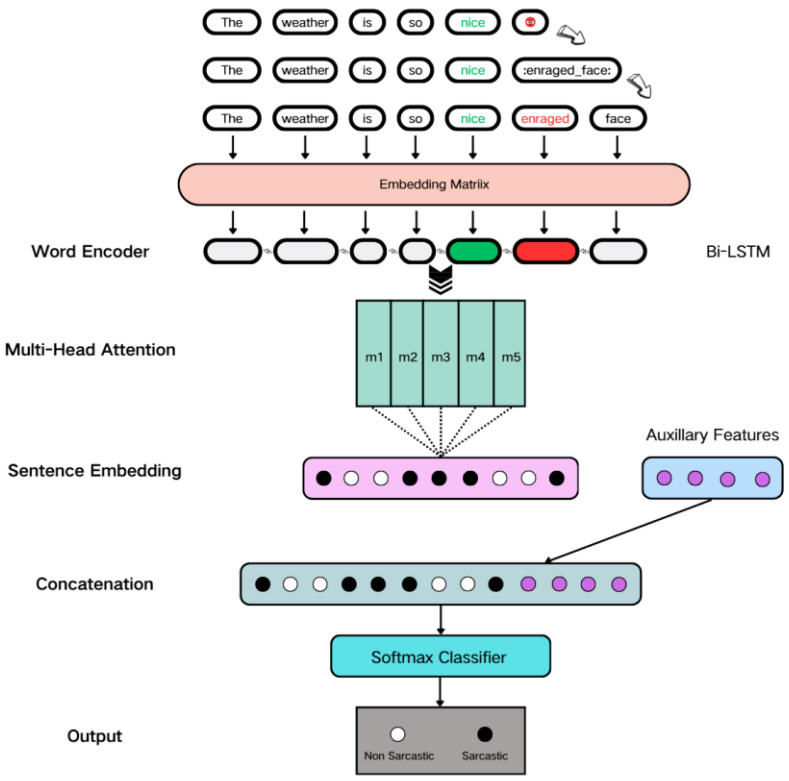

Fig. 2. Bidirectional LSTM model framework based on multi-head attention **(**Photo/Picture credit : Original)

These refinements ensure that the MHA-BiLSTM model achieves enhanced performance by facilitating a deeper and more nuanced understanding of the data structure, thereby advancing the state of the art in neural network design for complex sequence analysis tasks.

3 Dataset

This study focuses on the pivotal role of dataset selection in sarcasm detection. This paper employs the Self-Annotated Reddit Corpus (SARC), initially proposed by Khodak et al., as our primary dataset [16]. This extensive corpus includes over one million entries, equally divided between sarcastic and non-sarcastic comments sourced from the Reddit platform. Each entry is enriched with contextual details such as author profiles, ratings, and preceding comments, which are crucial for in-depth analysis. Within the framework of Reddit, users participate in topic-specific communities known as subreddits. The discussions are organized around initial posts termed "submits," where subsequent comments are structured in a hierarchical manner, ensuring each comment is linked to its predecessor termed a "parent comment." This format aids in preserving the conversational context, a key aspect in understanding sarcasm. Notably, SARC's reliability is bolstered by its method of annotation; users self-tag sarcastic remarks with the "/s" notation. This self-annotation approach significantly enhances the authenticity and accuracy of the data. The current iteration of the dataset comprises 505,413 sarcastic and an equal number of non-sarcastic statements, totaling 1,010,826 training samples. For evaluation purposes, the dataset also includes a balanced test set consisting of 251,608 comments. This structure provides a robust base for training and evaluating models aimed at sarcasm detection, thereby reflecting the dataset's importance in achieving reliable performance.

4 Results

4.1 Previous Works

In the examination, this paper benchmarked two models against advanced networks and foundational methods referenced by Hazarika et al. using a well-balanced SARC dataset variant [17]:

SVM: This model stands as a fundamental machine learning technique for sarcasm detection. It provides a comparative baseline which enables the assessment of more sophisticated approaches.

CNN-SVM: Crafted by Poria et al., this approach integrates a CNN to analyze textual content and another pretrained CNN to derive sentiment, emotion, and personality features. These features are then amalgamated and submitted to an SVM for the final classification step.

BiLSTM: This model excels in processing long-term dependencies, proving advantageous for detecting sarcasm in sentences due to its deep sequential learning capability.

CASCADE: Introduced by Hazarika et al., this technique incorporates user embeddings to capture individual personality and stylistic elements. It synergistically pairs these with a CNN to parse content features. This model presents outcomes from both the original and a modified version sans personality traits, underscoring the efficacy of our model even in the absence of such data.

These evaluations highlight not only the diverse methodologies applied in sarcasm detection but also underscore our efforts to refine detection through nuanced model configurations.

4.2 Results and Discussion

Table 1. Experimental results based on the SARC

Models	**Precision**	**Recall**	**F-Score**
No personality features			
SVM	72.30%	75.97%	74.09%
BiLSTM	69.41%	77.75%	73.34%
BERT + Intermediate Task Transfer Learning	**72.54%**	**83.51%**	**77.53%**
MHA-BiLSTM	**72.43%**	**82.73%**	**77.48%**
CASCADE(no personality features)	63.24%	70.13%	66.00%
With personality			
CNN-SVM	65.37%	71.43%	68.00%
CASCADE(with personality features)	72.18%	82.62%	77.00%

Table 1 presents the F-scores for the two evaluated models discussed in this paper, as well as results from prior research. The table is split into two sections: the first focuses solely on experiments conducted using sentence data without extra information, while the second part examines models that utilize the personality traits of the authors of the reviews. As the models reviewed in this study do not incorporate author information, their results are displayed in the first section. Here, it is evident that the models discussed outperform previous studies by a minimum margin of 3.0%, demonstrating their superior ability to grasp semantic nuances in sarcasm detection tasks.

In terms of both Precision and Recall, the BERT model based on transfer learning performs better than the pure Bidirectional LSTM based on Multi-Head Attention, especially in the accuracy of identifying sarcastic language. Its strength lies in its extensive language understanding ability obtained through large-scale corpus pre-training. Therefore, when capturing complex language phenomena such as sarcasm, it can more accurately understand the context and implicit meaning.

5 Conclusion

Sarcasm represents a sophisticated concept that often poses interpretative challenges, even for humans. This paper demonstrates the effectiveness of using a large pre-trained BERT language model and a bidirectional LSTM model based on a multi-head attention mechanism to accurately predict sarcasm. For the former, this model employs a task-specific transfer learning approach to refine the BERT model, optimizing its effectiveness. For the latter, this model integrates the multi-head attention with the bidirectional LSTM, boosting its capability to process sequential data accurately.

This paper finds that the BERT pre-trained language model based on transfer learning has significant advantages. Through large-scale corpus pre-training, it can comprehensively capture language knowledge and complex language phenomena, and simultaneously consider the left and right contextual information of each word in the sentence through the bidirectional Transformer architecture. In addition, the multi-head attention mechanism enables it to process different parts of a sentence and generate a complete feature

representation. Although this model combining LSTM and multi-head attention performs well in processing the contextual relationship of sequence data, it may not be as efficient as BERT in processing tasks with very complex and changing contexts. Especially on sarcasm recognition datasets such as SARC, which require high sensitivity to subtle differences in sarcasm.

In the future, the models reviewed in this paper can serve as a strong baseline for new research on this task, future work could use contextual data (e.g. user embeddings) to enhance the models can achieve new state-of-the-art performance. This paper hopes to make its decision-making process easier to understand and analyze by improving the interpretability of the model. Expanding the model to a multilingual environment and studying the universality and differences of sarcasm detection in different cultural backgrounds can be considered. Finally, it is suggested to plan to optimize the operating efficiency of the model through technologies such as model compression and distillation, so that the model can be effectively applied in resource-constrained scenarios.

Reference

1. J. Kao, E. Tong, V. Niculae, & C. Danescu-Niculescu-Mizil, SARC: A Corpus for Sarcasm Detection on Reddit. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, (2018).
2. E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, R. Huang, Sarcasm as contrast between a positive sentiment and negative situation, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, (2013).
3. S.K. Bharti, K.S. Babu, S.K. Jena, Parsing-based sarcasm sentiment recognition in twitter data, in: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social, (2015).
4. D.I.H. Farías, V. Patti, P. Rosso, Irony detection in twitter: The role of affective content, ACM Transactions on Internet Technology (TOIT) 16(3) (2016).
5. A. Ghosh, T. Veale, Fracking sarcasm using neural network, in: Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis, (2016).
6. A. Ghosh, T. Veale, Magnets for sarcasm: making sarcasm detection timely, contextual and very personal, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, (2017).
7. D.G. Maynard, M.A. Greenwood, Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. Lrec 2014 proceedings. ELRA, (2014).
8. Y. Ren, D. Ji, H. Ren, Context-augmented convolutional neural networks for twitter sarcasm detection, Neurocomputing 308, 1-7 (2018).
9. Y. Cai, H. Cai, X. Wan, Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July 2019–2 August (2019).
10. L. Li, O. Levi, P. Hosseini, D. Broniatowski, A Multi-Modal Method for Satire Detection using Textual and Visual Cues. In Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, Barcelona, Spain, 20 December (2020).
11. X. Wang, X. Sun, T. Yang, H. Wang, Building a Bridge: A Method for Image-Text Sarcasm Detection Without Pretraining on Image-Text Data. In Proceedings of the

First International Workshop on Natural Language Processing Beyond Text, Online, 20 November (2020).

12. Q. Zhang, J. Du, R. Xu, Sarcasm detection based on adversarial learning, Beijing Da Xue Xue Bao 55 (1) 29–36 (2019).
13. E. Savini, C. Caragea, Intermediate-Task Transfer Learning with BERT for Sarcasm Detection. *Mathematics, 10*, 844 (2022).
14. A. Kumar, V. T. Narapareddy, V. Aditya Srikanth, A. Malapati and L. B. M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM," in *IEEE Access*, vol. 8, pp. 6388-6397, 2020, doi: 10.1109/ACCESS. (2019).2963630.
15. Y. Cai, H. Cai, X. Wan, Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July 2019–2 August (2019).
16. M. Khodak, N. Saunshi, K. Vodrahalli, A Large Self-Annotated Corpus for Sarcasm. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May, (2018).
17. D. Hazarika, S. Poria, S. Gorantla, E. Cambria, R. Zimmermann, R. Mihalcea, CASCADE: Contextual Sarcasm Detection in Online Discussion Forums. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August (2018).