Proceedings of the 4th International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/79/2025.19930

Research Advanced in Sarcastic Detection Based on Deep Learning

Changsheng Shi[1,a,*]

[1]Warren College, University of California San Diego, CA, USA
a. c5shi@ucsd.edu
**corresponding author*

Abstract: Sarcasm detection is an important task in natural language processing (NLP), which aims to identify ironic intentions in text or discourse. Unlike conventional narratives, satire often uses descriptions that are opposite to the literal meaning to convey humor or criticism, and it is widely used in various scenarios such as social media, comment sections, and forums. Accurately detecting satire is crucial for understanding users' true intentions, while remains an open issue due to the freedom of language expression. Compared to face-to-face communication with the extra tone, facial expressions, and body movements, or an article with background and contextual information, a single satirical comment can hardly provide information for the machine to make a correct judgment, which may mislead the model to recognize the emotions contained in the comment. This article aims to introduce the role of different features in machine learning and deep learning satire detection tasks based on current research in this field. In addition, this article discusses the problems of satire detection and looks forward to its future development direction.

Keywords: Sarcasm detection, natural language processing, deep learning.

1. Introduction

Sarcasm is a type of irony used to express disapproval in a sharp or cutting manner, which often conveys an opposite meaning to its literal expression [1]. Due to its complexity, many of the traditional machine models cannot accurately detect sarcasm, making it one of the most challenging features for sentiment analysis in natural language processing. For instance, sarcasm can lead the algorithm to make a false judgment on the sentiment score. Such as comment like "Great job, you really nailed it!" might confuse the traditional model to give a positive score [2]. Although some works have employed advanced deep learning techniques in recent years, they still have limitations in recognizing subtle contextual cues.

Detection of sarcasm with high accuracy allows opinion mining, which can be beneficial for recommender systems based on reviews and conversational AI systems like those of Amazon's Alexa and Google Assistant, because the intent behind the user's query is important to recognize [3]. There is an increasing demand for systems that can find sarcasm from both what is written and spoken through social media and conversational agents [4]. Focusing on the sarcasm detection task, this paper provides a detailed investigation of the evolution of advanced technologies, from early approaches using machine learning to the state-of-the-art multimodal systems that include both text and audio.

2. Method

2.1. Approaches based on Syntactic Markers

Most of the early sarcasm detection methods were based on manual feature extraction techniques and hence relied heavily on lexical and syntactic markers. Linguistic features are crucial for this approach, such as patterns in punctuation, word frequency, and certain syntactic structures including exaggerated words or tonal cues [5]. One of notable traditional method is hashtag-based supervision, with being used usually to label sarcastic comments mainly from social media platforms like Twitter. These methods are simple in nature, lacked depth to get the whole context about sarcasm and has limitation of precision. One of the papers, titled "Sarcasm Detection in Product Reviews using Textual Entailment," developed the efficiency of hashtag-based supervision along with syntactic analysis. This was further found to be noisy due to reliance on user-generated hashtags that usually don't have any consistency in labeling sarcasm.

2.2. Approaches based on Machine Learning

The next step involved the application of machine learning models which included Support Vector Machines or SVMs, logistic regression, and random forests. These are early models that rely on hand-designed features such as word frequency, punctuation, and n-grams, to detect patterns in sarcastic comments [6]. Although these methods are more advanced compared to the traditional approach, problems persist in the identification of deeper, context-driven nuances of sarcasm. Šandor and Bagić Babac conducted a study, "Sarcasm Detection in Online Comments Using Machine Learning," in which different machine-learning models were compared. These included logistic regression, ridge, and SVMs. These models were trained on 1.3 million social media comments, abs showed a better performance using deep learning models.

2.3. Approaches based on Deep Learning

Models such as BiLSTM and BERT have revolutionized sarcasm detection in recent years. These models inherently extract features that effectively represent the subtleties of sarcastic remarks contextually, therefore improving accuracy without the need for exhaustive manual feature engineering [7]. Among them, BERT has turned out to be very effective in various natural language processing tasks due to its ability in capturing bidirectional dependencies in text. In the paper "Sarcasm Detection of Online Comments Using Emotion Detection," Shubham Rendalkar and Chaitali Chandankhede implemented a hybrid deep learning approach by fusing emotion detection along with sarcasm detection. It relies on the use of BERT and other lexical databases, such as WordNet and SentiWordNet, for assigning scores at an individual word level to a comment. Even though models like BERT have high computational costs, they outperform conventional machine learning approaches in the task of sarcasm detection significantly. In addition, Šandor and Bagić Babac have also shown that the BERT-based models outperformed other models such as SVMs and BiLSTM with an accuracy of 73.1% in the detection of sarcasm. This study indicates that how important deep learning is in performing the sarcasm detection task.

3. Experiment

3.1. Common datasets and metrics

The commonly used dataset in sarcasm detection is proposed by Šandor and Bagić Babac, which contains 1.3 million comments collected from Reddit made by Khodak et al. This dataset includes both sarcastic and non-sarcastic comments, which were identified using a sarcasm tag (\s) added by

Reddit users. This dataset is rich in informal language, sarcasm, and emotive content since it is directly collected from social media which is suitable for model testing. Šandor and Bagić Babac also did the following work for the data preprocessing: (1) Text Cleaning: Noise such as emoticons, punctuation, and special characters were removed from the comment text. Additionally, abbreviations were expanded to their full form (e.g., "won't" became "will not"). (2) Stop Word Removal: Common words that do not contribute much to the meaning of a sentence (like "the", "and" "is") were removed. (3) Tokenization: The text was split into individual words for easier processing. (4) Lowercasing and Lemmatization: All words were converted to lowercase, and lemmatization was applied to reduce words to their base. (5) Feature Extraction: Several features were manually extracted from the comments. (6) Sentiment Analysis: Each word was assigned a sentiment score using the VADER sentiment analysis tool. (7) Punctuation Count: The number of punctuation marks was calculated for each comment, as sarcasm is often expressed with exaggerated punctuation. (8) Uppercase Word Count: Sarcastic comments sometimes use all caps to express emphasis. (9) Polarity Difference: The sentiment difference between the comment and its parent comment was computed, as sarcasm often presents a sharp contrast in sentiment.

Accuracy, F1-score, precision, and recall will be used as the performance indicator. These metrics are important in determining each model's performance in finding sarcasm in text data.

3.2. Performance comparison

The performance of representative sarcasm detection methods are shown in Table 1. Traditional machine learning models, such as logistic regression, ridge regression and SVMs provide a baseline in the performance for sarcasm detection. The logistic regression method obtains an accuracy of 63.2%, F1-score of 60.4%, recall of 56.3% and precision of 65.4%. Ridge regression and SVM achieve a similar result, while is higher than the traditional logistic regression. While some of the basic models are effective, they are bound by their dependence on manually engineered features and restricted contextual understanding of sarcasm, thus accounting for the lower performance metrics than those obtained by deep learning models.

Most of the time, deep learning models such as BiLSTM and BERT outperform classic models by leveraging automatic feature extraction and an ability to know subtle nuances in contexts. While BiLSTM based models achieved an accuracy of 67%, precision of 68%, a F1-score roughly around 67.7% and recall of 68.3%, BERT model reaches a new record, accuracy of 73.1%, F1-score of 72.4%, recall of 71.3%, and precision of 72.2%. Since BERT captures the context from both directions, it can better recognize those subtle patterns in text that carry sarcasm. Therefore, at present, the best model to be relied on is BERT.

Table 1: Performance comparison of various representative methods

Method	Accuracy	Recall	Precision	F1-score
Logistic Regression	63.2%	56.3%	65.4%	60.4%
Ridge Regression	70.0%	65.6%	71.4%	68.4%
SVM	70.0%	65.8%	71.3%	68.3%
BiLSTM	67.0%	68.3%	68%	67.7%
BERT	73.1%	71.3%	72.2%	72.4%

4. Discussion

4.1. Strengths and Limitations

Compared to modern models, traditional ones have much less computational cost and are more interpretable, such as logistic regression and SVM. They do not need gigantic datasets or costly computational resources to perform; they are suitable for smaller datasets or resource-constrained environments. However, these models are heavily reliant on hand-crafted feature engineering, such as word frequency and n-grams, a process that is time-consuming and generally fails to capture the deeper, context-driven characteristics of sarcasm. As a result, their accuracies and F1-scores can be lower than those reached by the deep learning models. Ridge regression is surprisingly competitive compared to other more complex models, yielding competitive results of 70% accuracy with less feature engineering. These models constitute a sweet spot between performance and computational efficiency. While performing well, the challenges of machine learning models lie in detecting sarcasm, which relies on deeper semantic and contextual clues, hence limiting their potential.

BERT has outperformed the rest of the other classic models, which is benefited from the bidirectional context and fine-tuning capabilities. BiLSTM, on the other hand, does not provide and ideal outcomes, even though it also automatically extracts features and captures complex patterns from raw text. The potential reason may be overfitting. By randomly selecting 10 thousand comments from the dataset and splitting it into 8:2 ration, the accuracy of the training dataset can reach 91.77%, while is only 56.71% on the testing dataset. This shows that the provided BilSTM model is overfitting. The major disadvantage of the deep learning models, especially that of BERT, is the computational cost. These models must be trained on huge amounts of data and need considerable computational resources, which in some applications is not feasible. Moreover, BERT is more difficult to interpret than lighter machine learning models due to its complexity [8].

4.2. Future direction

Current research is predominantly concentrated on English language datasets, while the detection of sarcasm for a multilingual data set has much less study. Researchers like SarcasmDet in Arabic show how important it is to broaden and make models more inclusive. Another direction in the future is that many models that have been trained on specific datasets tend to fail on new domains; for instance, models moving from tweet data to news headlines [9]. Instead, through transfer learning, domain adaptation techniques, the model will be able to generalize among multiple contexts.

Social media and real-time data analysis are common sources of satirical language. With the continuous growth of social media platforms, the ability to perform satire detection on real-time data streams will become particularly important. The advancement of real-time processing technology will enable the system to respond faster to newly emerging content. Meanwhile, with the growth of multimedia data, future satire detection systems will rely more on multimodal information, including not only text but also audio and video data. For example, in videos, facial expressions, tone of voice, and body language can all provide important clues about whether the speaker is using satire. This relies on the optimization of deep learning models, especially Transformers, to make them more efficient and robust, while reducing the demand for computing resources. The understanding of context aware satire is highly dependent on the context [10]. Future satire detection systems will place greater emphasis on capturing contextual information, including factors such as context, speaker identity, and audience response.

5. Conclusion

This paper focuses on the advanced technologies in sarcasm detection, which details the representative methods from the three aspects. Sarcasm detection has evolved in the literature from straightforward, rule-based approaches to more recent machine learning and deep learning-based techniques. While earlier approaches using lexical analysis or supervision based on hashtags provided insight into the problem, they were not subtle enough for complex sarcastic comments. Then, machine learning models managed to improve these with feature engineering but again failed to capture contextual depth in sarcasm. Deep learning models, particularly BERT, have demonstrated high performance by unanimously extracting features, keeping in consideration bidirectional dependencies.

References

[1] *Filik, R., Ţurcan, A., Ralph-Nearman, C., & Pitiot, A. (2019). What is the difference between irony and sarcasm? An fMRI study. Cortex, 115, 112–122.*

[2] *Yacoub, A. D., Slim, S. O., & Aboutabl, A. E. (2024). A survey of sentiment analysis and sarcasm detection: Challenges, techniques, and trends. International Journal of Electrical and Computer Engineering Systems, 15(1), 69-78.*

[3] *Šandor, D., & Babac, M. B. (2023). Sarcasm detection in online comments using machine learning. Information Discovery and Delivery, 52(2), 213–226. https://doi.org/10.1108/idd-01-2023-0002*

[4] *Sobti, S., Agarwal, V., Tiwary, A., Naval, P., Jayabalan, B., & Sohal, J. (2024). Sarcasm in the digital age: An opinion. Nanotechnology Perceptions, 20(S3), 944-959.*

[5] *Sinha, S., Vijeta, T., Kubde, P., Gajbhiye, A., Radke, M. A., & Jones, C. B. (2023). Sarcasm detection in product reviews using textual entailment approach. Proceedings of the 2023 7th International Conference on Natural Language Processing and Information Retrieval (NLPIR 2023), Seoul, Republic of Korea, 310-318.*

[6] *Bharti, S. K. (2019). Sarcasm detection in textual data: A supervised approach (Doctoral dissertation, National Institute of Technology Rourkela). http://ethesis.nitrkl.ac.in/10002/*

[7] *Baruah, A., Das, K. A., Barbhuiya, F. A., & Dey, K. (2020). Context-aware sarcasm detection using BERT. In Proceedings of the Second Workshop on Figurative Language Processing (pp. 83-87). https://doi.org/10.18653/v1/2020.figlang-1.12*

[8] *Guo, X., Li, B., Yu, H., & Miao, C. (2021). Latent-optimized adversarial neural transfer for sarcasm detection. arXiv preprint arXiv:2104.09261. https://arxiv.org/abs/2104.09261*

[9] *Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., & Mihalcea, R. (2018). CASCADE: Contextual sarcasm detection in online discussion forums. Proceedings of the 2018 International Conference on Computational Linguistics. arXiv preprint arXiv:1805.06413.*

[10] *Doona, J. (2020). News satire engagement as a transgressive space for genre work. International Journal of Cultural Studies, 24(1), 15-33.*