

Unmasking Sarcasm: Sarcastic Language Detection with BiLSTMs

Anusha M D^{1,*}, Parameshwar R. Hegde^{1,†}

¹Department of Computer Science, Yenepoya Institute of Arts Science Commerce and Management, Yenepoya (Deemed to be University), Balmata, Mangalore

Abstract

Across the globe, there's a clear and growing tendency to incorporate sarcasm into everyday interactions. This rising trend can be primarily attributed to the widespread adoption of sarcasm in daily life, with a particular emphasis on its prevalence on social media and the Internet. Sarcasm poses a significant challenge for sentiment analysis systems as it communicates opinions indirectly, often deviating from literal meanings. There is a growing demand for detecting sarcasm and sentiment in code-mixed social media content written in Dravidian languages. This paper describes the Bidirectional Long Short Term Memory (BiLSTM) model and obtained weighted F1-scores of 0.58 and 0.63 for the Ta-En and Ma-En pairs respectively. These results have been submitted to "Sarcasm Identification of Dravidian Code-Mixed@FIRE-2023" to analyze the sarcasm in Malayalam-English (Ma-En) and Tamil-English (Ta-En) code-mixed texts.

Keywords

Bi-LSTM, Code-Mixed, Deep Learning, Dravidian Languages, Sarcasm

1. Introduction

The rise of social media over the past ten years has united all regions of the world into a central hub for enhancing communication [1]. One of the most popular channels for people to share information and voice their opinions is social media. These statistics are used by many governments and corporations to measure public opinion on various goods, entertainment options, and political issues. Sarcasm represents a sophisticated way of conveying emotions, in which the speaker expresses their opinions opposite of what they mean. Sarcasm is often characterized as ironic or satirical which is used to insult, mock, or amuse. Sarcasm takes on various definitions in different dictionaries based on their perspectives [2]. As outlined in the Macmillan English dictionary¹, sarcasm involves "expressing the opposite of one's true meaning verbally or in writing, often with the intention of making someone appear foolish or revealing anger." According to The Random House dictionary², sarcasm entails "biting or harsh

Woodstock'23: Symposium on the irreproducible science, December 15-18, 2023, Woodstock, NY

*Corresponding author.

†These authors contributed equally.

✉ anushamd@yenepoya.edu.in (A. M. D); parameshwarhegde@yenepoya.edu.in (P. R. Hegde)

🆔 0009-0000-3644-1260 (A. M. D)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹<http://www.macmillandictionary.com/>

²<http://www.thefreedictionary.com/>

mockery or irony” and ”a cutting taunt or scornful comment.” The Collins English dictionary ³ characterizes it as ”language that mocks, expresses contempt, or uses irony to convey insults or disdain.” Merriam-Webster ⁴ presents another interpretation, describing sarcasm as ”a form of satirical wit that relies on acerbic, biting, and frequently ironic language, typically aimed at an individual.

Several online social networking platforms enable users to share and access posts expressing their views on various subjects such as products, politics, stock markets, and entertainment [3]. Additionally, users often compose messages with intricate sentence structures, posing challenges for both machines and humans to grasp their intended meaning. Consequently, sentiment analysis, along with its practical application like sarcasm detection, has emerged as a prominent trend in the field of data mining.

Based on the analysis of Pranali Chaudhari [3] studies, it can be inferred that there are several distinctive features in text that play a significant role in the identification of sarcasm. These features include lexical elements, hyperbole, and pragmatic cues[4]. Various methodologies have been employed to detect sarcasm. In recent years, a majority of research has leveraged supervised or semi-supervised machine learning approaches for this purpose [5][6] [3]. Additionally, novel strategies [7],[8], behavioral approaches [9], and bootstrapping techniques [10] have been applied in sarcasm detection.

On other hand, Traditional investigations, such as those by Davidov et al. [11] and Riloff et al. [10], employed rule-based methods to address sarcasm detection. However, more recent research [12] has shifted towards the utilization of deep learning techniques to automatically detect the discriminatory features. In this study, a deep learning-based Bidirectional Long Short-Term Memory (BiLSTM) Neural network(NN) is employed, which consists of LSTM units that integrate past and future context information because of which they are showing excellent performance for sequential modelling problems as well as for Text Classification (TC) [13]. NN models expect numeric values as input. Hence, it is necessary to convert the text data to numeric representation by building an embedding layer before building a BiLSTM model.

2. Literature Review

Sarcasm identification task has been studied by employing different methods, including lexicon-based, machine learning, deep learning, or even a hybrid approach. Some of the research works related to Sarcasm detection are discussed below.

Ibrahim Abu-Farha et.al [14] introduced ArSarcasm, a dataset designed for detecting sarcasm in the Arabic language. This dataset was constructed by reevaluating existing Arabic sentiment analysis datasets, resulting in a collection of 10,547 tweets, with approximately 16% of them being identified as containing sarcasm. Beyond sarcasm, the dataset also underwent annotation for sentiment and dialect characteristics. The analysis highlights the highly subjective nature of these tasks, as indicated by the varying sentiment labels influenced by annotators’ biases. Experiments demonstrate the limitations of state-of-the-art sentiment analyzers when confronted

³<https://www.collinsdictionary.com/>

⁴<http://www.merriam-webster.com/>

with sarcastic content. Finally, a deep learning model using BiLSTM was trained for sarcasm detection and the model achieved an F1-score of 0.46.

Priya Goel et.al [15] aims to narrow the gap between human and machine intelligence when it comes to recognizing and comprehending sarcasm in online behavior and patterns. The study utilizes neural techniques such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and a Baseline Convolutional Neural Network (CNN) in an ensemble model designed to detect sarcasm on the internet. The study used the news Headlines dataset and for the Reddit dataset and to enhance the accuracy of the proposed model, the necessary dataset is prepared using pre-trained word embedding models such as fastText, Word2Vec, and GloVe, and their performance is compared. The objective is to assess the overall sentiment of the writer as either positive or negative, and whether the text is sarcastic or not, to ensure the intended message is correctly conveyed to the audience. The final findings indicate that the proposed ensemble model, when combined with word embeddings, outperformed other state-of-the-art models and deep learning models considered in the study. It achieved an accuracy rate of approximately 96% for the News Headlines dataset, 73% for the Reddit dataset, and among proposed ensemble models, the Weighted Average Ensemble obtained the highest accuracy at around 99% and 82% for both datasets, respectively. The use of an ensemble model significantly improved the stability, precision, and predictive capabilities of the proposed approach.

Abdelkader El Mahdaouy et.al[16] proposed an end-to-end deep Multi-Task Learning model for Sentiment Analysis(SA) and sarcasm detection. The ArSarcasm Shared Task consists of two subtasks for sarcasm detection and SA in Arabic language. This study leverages the MARBERT's contextualized word embedding with a multi-task attention interaction module.The MTL model's architecture consists of a Bidirectional Encoder Representation from Transformers (BERT) model, a multi-task attention interaction module, and two task classifiers. The aim is to allow task interaction and knowledge sharing for both SA and sarcasm detection. The model shows very promising results on both subtasks.

Mudoor Devadas Anusha et al [13] proposed Bidirectional Long Short Term Memory (BiLSTM) model submitted to "Sentiment Analysis of Dravidian Languages in Code-Mixed Text" to analyze the sentiments in Kannada-English (Kn-En), Malayalam-English (Ma-En), and Tamil-English (Ta-En) code-mixed texts. In the proposed approach, the code-mixed word embeddings' are constructed using the training set of the respective code-mixed language pairs' and these embeddings are used to build a Deep Learning (DL) model based on BiLSTM. The proposed model obtained weighted F1-scores of 0.563, 0.604, and 0.365 for code-mixed Ta-En, Ma-En and Kn-En language pairs respectively.

C. I. Eke et. al [17] introduces a context-driven feature methodology for sarcasm identification, employing a combination of the BERT model, deep learning, and conventional machine learning to tackle the aforementioned challenges. It leverages two benchmark datasets, namely the Twitter and Internet Argument Corpus, version two (IAC-v2), for classification across the three learning models. The first model employs an embedding-based approach within a deep learning framework, utilizing Bidirectional Long Short-Term Memory (Bi-LSTM), a variant of Recurrent Neural Network (RNN), and Global Vector representation (GloVe) for word embedding and context comprehension. The second model is founded on Transformer architecture, employing a pre-trained Bidirectional Encoder representation and Transformer (BERT). In contrast, the third model adopts feature fusion, incorporating BERT features, sentiment-related features, syntactic

features, and GloVe embedding features within a conventional machine learning framework. To assess the effectiveness of this methodology, comprehensive evaluation experiments were conducted. Remarkably, when applied to two Twitter benchmark datasets, the technique achieved the highest precision of 98.5% and 98.0%, respectively.

3. Methodology

3.1. Pre-processing

To enhance the classifier's performance, it is essential to undertake data preprocessing, aimed at eliminating any unwanted noise. Text pre-processing procedures may differ depending on the task and the dataset used. The following the preprocessing steps employed in our proposed approach.

- Converting all text to lowercase, as character case is irrelevant for TC.
- Removing numeric and punctuation characters, as they hold no importance for TC.
- Label encoding, which converts class/category labels into numerical values to make them machine-readable.

3.2. Feature Engineering

Next phase is converting Pre-processing text into vectors in this proposed approach, First we created a dictionary where words from the text are mapped to their respective word embeddings (vectors). Then word's vector representation from the 'ftw2v' model which used as pre-trained word embedding model. dimension of word embeddings. It's set to 300, indicating that each word will be represented as a vector in a 300-dimensional space.

3.3. Model Construction

BiLSTM is the neural network model for text classification, including an embedding layer, bidirectional LSTM layer, and output layer. The model is initialized as a sequential neural network, which is a linear stack of layers. Then Embedding layer converts integer-encoded words into dense vectors of fixed size. These vectors are trainable during model training and capture semantic relationships between words. Spatial dropout helps prevent overfitting by randomly setting a fraction of input units to zero at each update during training. The benefit of Bidirectional LSTM is process sequences in both directions, capturing information from past and future contexts. This can be particularly useful for understanding the context of words in a sentence. sigmoid activation is used for binary classification tasks. It assigns probabilities to each class, and the class with the highest probability is predicted as the output. Output dimensions of the model are configured based on the number of class labels. The structure of the BiLSTM model is shown in Figure 1. Overall, this neural network architecture is designed to effectively process and classify text data into two classes

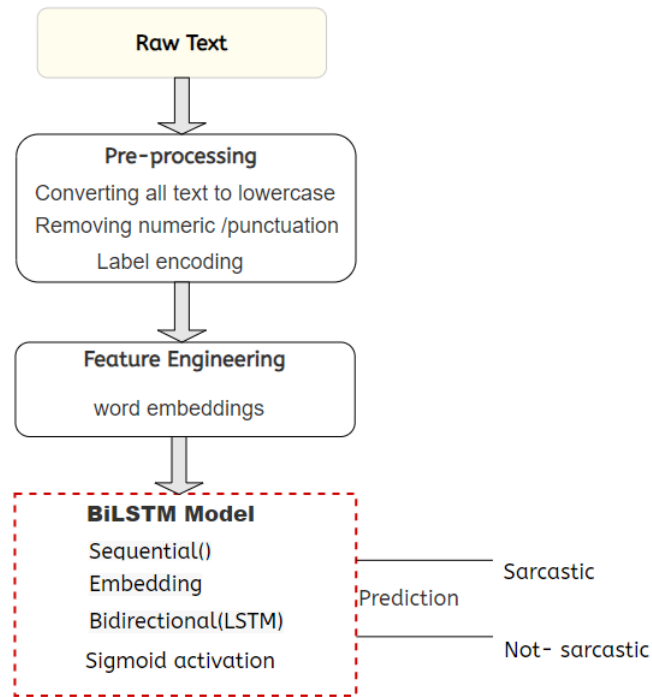


Figure 1: Structure of the BiLSTM Model

Table 1

Distribution of labels in the given dataset

Language-Pair	Labels	Train Set	Validate Set	Test Set
Malayalam-English	Sarcastic	2259	588	2263
	Non-Sarcastic	9798	6186	2427
Tamil- English	Sarcastic	7170	1820	2264
	Non-Sarcastic	19866	4939	6186

Table 2

Results of the proposed models on the Development sets

Language-Pair	Precision	Recall	F1-score
Malayalam-English	0.83	0.81	0.82
Tamil- English	0.78	0.77	0.77

4. Experimental setup and Results

The dataset⁵[18] provided by the organizers of the Sarcasm Identification task for Dravidian Languages Malayalam and Tamil in DravidianCodeMix contains Train , Development, and Test

⁵<https://codalab.lisn.upsaclay.fr/competitions/13540participate-get-data>

sets. This task involves Tamil-English and Malayalam-English datasets containing YouTube video comments. The comments in the dataset are predominantly composed in both the native script and Roman script, featuring either Tamil/Malayalam grammar alongside English vocabulary or English grammar combined with Tamil/Malayalam vocabulary. Table 1 shows the distribution of labels in given dataset. Additionally, it is noteworthy that the training dataset exhibits a significant class imbalance, as depicted in graphical representations provided in Figures 2 and 3, which may impact the model’s performance and result interpretation.

Scikit-learn⁶ and Keras⁷ which are minimalist libraries for deep learning, were employed for implementing the Python code. A BiLSTM model with word embedding features was applied to the Test set of all three language pairs, The model obtained weighted F1-scores of 0.68 and 0.63 for the Ta-En and Ma-En pairs respectively. The outcomes of our experiments on the test dataset have been made available on the “Sarcasm Identification” task page⁸. Furthermore, Table 2 displays the performance of the proposed approach on the Development sets of the Ta-En and Ma-En language pairs

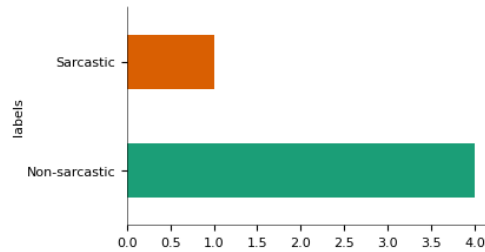


Figure 2: Imbalance Distribution of Ta-En Train set

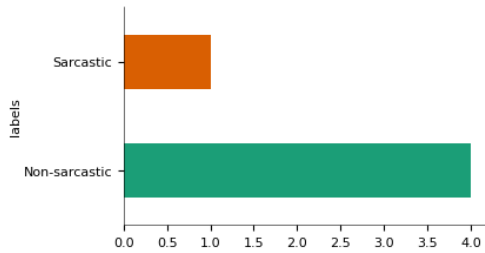


Figure 3: Imbalance Distribution of Ma-En Train set

5. Conclusion

In this research presents the details of proposed operational model designed for detecting sarcasm in code-mixed text written in Malayalam and Tamil. The results of this model has been submitted to “Sarcasm Identification of Dravidian Languages (Malayalam and Tamil) in DravidianCodeMix” a shared task organized by DravidianLangTech. To address the challenge of categorizing YouTube video comments into predefined categories, this study introduce a

⁶<https://scikit-learn.org/>

⁷<https://keras.io/api/layers/>

⁸https://codalab.lisn.upsaclay.fr/competitions/13540learn_he_details-results

BiLSTM model that utilizes word embeddings as its features. The proposed model achieved F1-scores of 0.563 and 0.604 for the Ta-En and Ma-En language pairs respectively. The future work aim to increase existing dataset and leverage advanced technologies to enhance the model's performance.

References

- [1] B. R. Chakravarthi, A. Hande, R. Ponnusamy, P. K. Kumaresan, R. Priyadharshini, How can we detect homophobia and transphobia? experiments in a multilingual code-mixed setting for social media governance, *International Journal of Information Management Data Insights* 2 (2022) 100119.
- [2] B. R. Chakravarthi, N. Sripriya, B. Bharathi, K. Nandhini, S. Chinnaudayar Navaneethakrishnan, T. Durairaj, R. Ponnusamy, P. K. Kumaresan, K. K. Ponnusamy, C. Rajkumar, Overview of the shared task on sarcasm identification of Dravidian languages (Malayalam and Tamil) in *DravidianCodeMix*, in: *Forum of Information Retrieval and Evaluation FIRE - 2023*, 2023.
- [3] P. Chaudhari, C. Chandankhede, Literature survey of sarcasm detection, in: *2017 International conference on wireless communications, signal processing and networking (WiSPNET)*, IEEE, 2017, pp. 2041–2046.
- [4] F. Kunneman, C. Liebrecht, M. Van Mulken, A. Van den Bosch, Signaling sarcasm: From hyperbole to hashtag, *Information Processing & Management* 51 (2015) 500–509.
- [5] E. Fersini, F. A. Pozzi, E. Messina, Detecting irony and sarcasm in microblogs: The role of expressive signals and ensemble classifiers, in: *2015 IEEE international conference on data science and advanced analytics (DSAA)*, IEEE, 2015, pp. 1–8.
- [6] D. Bamman, N. Smith, Contextualized sarcasm detection on twitter, in: *proceedings of the international AAAI conference on web and social media*, volume 9, 2015, pp. 574–577.
- [7] P. Liu, W. Chen, G. Ou, T. Wang, D. Yang, K. Lei, Sarcasm detection in social media based on imbalanced classification, in: *Web-Age Information Management: 15th International Conference, WAIM 2014, Macau, China, June 16-18, 2014. Proceedings* 15, Springer, 2014, pp. 459–471.
- [8] F. Barbieri, H. Saggion, F. Ronzano, Modelling sarcasm in twitter, a novel approach, in: *proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2014, pp. 50–58.
- [9] A. Rajadesingan, R. Zafarani, H. Liu, Sarcasm detection on twitter: A behavioral modeling approach, in: *Proceedings of the eighth ACM international conference on web search and data mining*, 2015, pp. 97–106.
- [10] 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, pp. 704–714.
- [11] D. Davidov, O. Tsur, A. Rappoport, Semi-supervised recognition of sarcasm in twitter and amazon, in: *Proceedings of the fourteenth conference on computational natural language learning*, 2010, pp. 107–116.

- [12] S. Poria, E. Cambria, D. Hazarika, P. Vij, A deeper look into sarcastic tweets using deep convolutional neural networks, arXiv preprint arXiv:1610.08815 (2016).
- [13] M. D. Anusha, H. L. Shashirekha, Bilstm-sentiments analysis in code-mixed dravidian languages (2021).
- [14] I. A. Farha, W. Magdy, From arabic sentiment analysis to sarcasm detection: The arsarcasm dataset, in: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, 2020, pp. 32–39.
- [15] P. Goel, R. Jain, A. Nayyar, S. Singhal, M. Srivastava, Sarcasm detection using deep learning and ensemble learning, Multimedia Tools and Applications 81 (2022) 43229–43252.
- [16] A. E. Mahdaouy, A. E. Mekki, K. Essefar, N. E. Mamoun, I. Berrada, A. Khoumsi, Deep multi-task model for sarcasm detection and sentiment analysis in arabic language, arXiv preprint arXiv:2106.12488 (2021).
- [17] C. I. Eke, A. A. Norman, L. Shuib, Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model, IEEE Access 9 (2021) 48501–48518.
- [18] B. R. Chakravarthi, Hope speech detection in youtube comments, Social Network Analysis and Mining 12 (2022) 75.