Sarcasm Detection in Dravidian Languages Using Machine Learning and Transformer Models

Malliga Subramanian, Aruna A, Anbarasan T, Amudhavan M, Kogilavani S V [1]

[1]Kongu Engineering College Erode Tamil Nadu India

Abstract

Sarcasm detection, particularly in code-mixed languages like Tamil-English and Malayalam-English, has become an increasingly important challenge in natural language processing (NLP) due to the growing use of social media. This paper presents various machine learning and transformer-based models, including Random Forest, Decision Trees, K-Nearest Neighbors, BERT, RoBERTa, and ALBERT, to detect sarcasm in Dravidian languages. We evaluate these models based on accuracy, precision, recall, and F1-score, using code-mixed datasets from social media platforms such as YouTube. Our study shows that transformer models, particularly RoBERTa, outperform traditional classifiers in detecting sarcasm. Future research aims to explore hybrid models and advanced pre-processing techniques

Keywords

Sarcasm Detection, Dravidian Language, code-mixed text, Machine Learning, Transformer models, BERT, RoBERTa

1. Introduction

Sarcasm, a linguistic tool where the intended meaning of a sentence differs from the literal meaning, presents a significant challenge for sentiment analysis. With the rise of social media platforms, there has been an increasing need to detect sarcasm automatically, especially in multilingual, code-mixed environments. In particular, sarcasm detection in Dravidian languages like Tamil and Malayalam, often intertwined with English, has become crucial for creating more effective sentiment analysis systems. The complexity of sarcasm detection is exacerbated when texts are code-mixed, i.e., when two or more languages are used interchangeably. Traditional sentiment analysis models fail to perform well in these scenarios as they are usually trained on monolingual datasets. This paper explores various approaches, including traditional machine learning models and transformer-based models like BERT and RoBERTa, to detect sarcasm in Dravidian languages.

2. Literature Survey

Sarcasm detection is an essential area in natural language processing (NLP), particularly for sentiment analysis, opinion mining, and emotion recognition. While substantial advancements have been made in sarcasm detection for English and other widely studied languages, research for low-resource languages, like Malayalam and Tamil (Dravidian language), remains limited. Sarcasm detection is challenging because it relies heavily on context, tone, and cultural nuances, making it a difficult task for machine learning and deep learning models. The survey provides an overview of the models submitted by participants for the task of sarcasm identification in Dravidian languages, as presented in DravidianCodeMix@FIRE-2024. It highlights the methodologies employed, the diversity of approaches, and the overall contributions of each submission, aiming to enhance understanding and improve future efforts in this area of research [14].

2.1. Sarcasm Detection in English and Major Languages

Initial efforts in sarcasm detection were primarily rule-based, utilizing lexical and syntactic analysis to identify specific patterns [1]. Davidov et al. (2010) introduced a semi-supervised technique using features such as patterns, punctuation, and n-grams from Twitter data. While effective in certain contexts, these approaches struggled with generalization due to sarcasm's context-dependent nature [3]. As machine learning advanced, supervised methods like Support Vector Machines (SVM), Random Forests (RF), and Logistic Regression (LR) became popular in sarcasm detection, often relying on hand-crafted features like n-grams, sentiment lexicons, and part-of-speech tags [5]. However, these methods faced limitations in capturing deeper linguistic nuances. With the introduction of transformer models like BERT and its variants (e.g., RoBERTa, ALBERT), significant improvements have been seen in sarcasm detection [2]. These models utilize self-attention mechanisms to understand word relationships in context, leading to state-of-the-art performance in English sarcasm detection.

2.2. Sarcasm Detection in Dravidian Languages

Although sarcasm detection research in English has made great strides, studies focusing on Dravidian languages such as Malayalam and Tamil remain sparse. These languages are morphologically rich and syntactically complex, which adds difficulty to sarcasm detection. Moreover, the lack of large annotated datasets poses another challenge. Most research in Malayalam and Tamil has concentrated on sentiment analysis and emotion detection using traditional machine learning models like Naive Bayes, SVM, and RF. The survey provides a comprehensive overview of the various models and techniques used by participants in the DravidianCodeMix@FIRE-2024 challenge to identify sarcasm in Dravidian languages [14]. The reliance on manually engineered features limits these models in fully capturing the complexity of sarcasm. Proper identification of non-sarcastic content is crucial for accurate sentiment analysis and reduces false positives in sarcasm detection [6][7].

2.3. Deep Learning and Transformers

Given the success of deep learning models in other languages, there is a growing interest in applying these techniques to sarcasm detection for Malayalam and Tamil. Models like CNNs, Multilayer Perceptron (MLP), and Gated Recurrent Units (GRU) have shown promise for text classification by capturing semantic and sequential information. However, their application to sarcasm detection in Malayalam and Tamil is still underexplored due to the lack of annotated data. Transformer models such as BERT, RoBERTa, and GPT have revolutionized NLP tasks by capturing complex relationships through self-attention mechanisms. These models could significantly enhance sarcasm detection in Malayalam and Tamil by leveraging transfer learning, where models pre-trained on large datasets, like multilingual BERT, are fine-tuned for Malayalam and Tamil specific tasks. Early research in sentiment analysis for Dravidian languages using transformer models shows potential for sarcasm detection, provided sufficient annotated data for fine-tuning is available.

2.4. Challenges

The main challenge in sarcasm detection for Malayalam and Tamil is the scarcity of large annotated datasets. Annotating sarcasm is labor-intensive due to its nuanced nature. Additionally, Malayalam's rich morphology and inflectional changes complicate text normalization and feature extraction. Future research should focus on building larger annotated corpora and developing hybrid approaches that combine traditional machine learning with deep learning architectures to better address sarcasm's complexity. Although sarcasm detection for Malayalam and Tamil and other Dravidian languages is still in its infancy, modern deep learning techniques and transfer learning approaches hold promise for improving the task.

3. Materials and Methods

The dataset used in this study consists of code-mixed comments in Tamil-English and Malayalam-English, collected from social media platforms like YouTube, Facebook, Twitter [12][13]. The dataset includes 6200 comments in the training set and 700 in the test set, with labels indicating whether the comment is sarcastic or not. The dataset reflects a real-world class imbalance, with more non-sarcastic comments than sarcastic ones. Techniques like SMOTE were used to balance the classes.

3.1. Preprocessing and Feature Extraction

Preprocessing is essential for managing the noisy nature of social media text, especially in sarcasm detection [5][11]. Key steps include removing special characters, emojis, and URLs to clean and standardize the text [10]. Another important step is transliteration, which converts Tamil and Malayalam text into a consistent script for easier processing. Tokenization and vectorization are also crucial, where techniques like CountVectorizer and TF-IDF are used to convert text into feature matrices. For instance, CountVectorizer transforms words into token counts, creating structured matrices that can be effectively used for classification tasks. These preprocessing steps ensure that the text is in a clean, structured format suitable for model training.

3.2. Models and Methodology

Several traditional machine learning models were employed for sarcasm detection, including Random Forest (RF), Decision Trees (DT), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). These models are widely used for classification tasks due to their ability to handle different types of data. In addition to these, transformer-based models have also been explored, such as BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (Robustly Optimized BERT), and ALBERT (A Lite BERT) [8]. These models provide deep contextual understanding, making them highly effective for detecting sarcasm in text.

4. Results and Discussion

The study on sarcasm detection in Tamil and Malayalam revealed that while traditional machine learning models like Random Forest and Support Vector Machine performed reasonably well, transformer-based models, particularly BERT and RoBERTa, significantly outperformed them due to their ability to grasp contextual nuances.

The models were evaluated using accuracy, precision, recall, and F1-score. Table 1 shows the performance of each model:

Table 1
Performance of Models for Sarcasm Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	76.12	72.45	78.56	75.36
Decision Trees	74.89	69.34	75.20	72.17
KNN	75.21	70.89	76.45	73.56
SVM	78.34	72.56	80.12	76.22
BERT	83.67	79.45	85.78	82.50
RoBERTa	86.89	82.34	88.45	85.30
ALBERT	82.12	78.56	83.89	81.12

RoBERTa outperformed other models in terms of accuracy and F1-score, highlighting the superiority of transformer-based models in detecting sarcasm in code-mixed text [4].

```
                        precision    recall  f1-score   support

      Non-sarcastic         0.85      0.95      0.89      4318
          Sarcastic         0.79      0.54      0.64      1596

           accuracy                             0.84      5914
          macro avg         0.82      0.74      0.77      5914
       weighted avg         0.83      0.84      0.83      5914
```

Figure 1: Classification report for Tamil

```
                    precision    recall  f1-score   support

   Non-sarcastic        0.87      0.99      0.93      2142
       Sarcastic        0.90      0.37      0.52       496

        accuracy                            0.87      2638
       macro avg        0.88      0.68      0.73      2638
    weighted avg        0.88      0.87      0.85      2638
```

Figure 2: Classification report for Malayalam

5. Conclusion

This study investigates sarcasm detection in code-mixed Dravidian languages, comparing various classifiers such as Naive Bayes Multinomial, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and the transformer-based model RoBERTa. The results revealed that RoBERTa significantly outperformed traditional models, demonstrating superior accuracy in distinguishing sarcastic from non-sarcastic content, thanks to its ability to leverage contextual embeddings and large training datasets. The analysis identified key linguistic features contributing to sarcasm, including idiomatic expressions and cultural references that vary among speakers. Future work will focus on exploring alternative feature representations, such as word embeddings, and developing hybrid models that combine traditional and deep learning approaches. This research aims to enhance sarcasm detection systems, improving their effectiveness in multilingual contexts, particularly in social media and conversational AI, where sarcasm is a prevalent form of communication [9].

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] M. Bouazizi, T. Ohtsuki, "A pattern-based approach for sarcasm detection on Twitter," *IEEE Access*, vol. 4, pp. 5477-5488, 2016. doi: 10.1109/ACCESS.2016.2598816.

[2] A. Agrawal, A. An, "Affective representations for sarcasm detection," in *Proceedings of the 41st International ACM SIGIR Conference on Research Development in Information Retrieval*, pp. 1029-1032, 2018.

[3] S. K. Bharti, K. S. Babu, R. Raman, "Context-based sarcasm detection in Hindi tweets," in *Proceedings of the 9th International Conference on Advances in Pattern Recognition (ICAPR)*, pp. 1-6, 2017. doi: 10.1109/ICAPR.2017.24.

[4] M. S. M. Suhaimin, M. H. A. Hijazi, R. Alfred, F. Coenen, "Natural language processing based features for sarcasm detection: An investigation using bilingual social media texts," in *Proceed-*

ings of the 8th International Conference on Information Technology (ICIT), pp. 703-709, 2017. doi: 10.1109/ICIT.2017.7976604.

[5] D. Das, A. J. Clark, "Sarcasm detection on Facebook: A supervised learning approach," in *Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct*, pp. 1-5, 2018.

[6] S. Gupta, R. Singh, V. Singla, "Emoticon and text sarcasm detection in sentiment analysis," in *Proceedings of the 1st International Conference on Sustainable Technologies for Computational Intelligence*, Springer, Singapore, pp. 1-10, 2020. doi: 10.1007/978-981-15-4294-01.

[7] M. J. Adarsh, P. Ravikumar, "Sarcasm detection in text data to bring out genuine sentiments for sentimental analysis," in *Proceedings of the 1st International Conference on Advances in Information Technology (ICAIT)*, pp. 94-98, 2019. doi: 10.1109/ICAIT.2019.00027.

[8] J. Lemmens, B. Burtenshaw, E. Lotfi, I. Markov, W. Daelemans, "Sarcasm detection using an ensemble approach," in *Proceedings of the Second Workshop on Figurative Language Processing*, pp. 264-269, 2020.

[9] Y. A. Kolchinski, C. Potts, "Representing social media users for sarcasm detection," *arXiv preprint arXiv:1808.08470*, 2018.

[10] M. Khodak, N. Saunshi, K. Vodrahalli, "A large self-annotated corpus for sarcasm," *arXiv preprint arXiv:1704.05579*, 2017.

[11] D. Das, A. J. Clark, "Sarcasm detection on Flickr using a CNN," in *Proceedings of the 2018 International Conference on Computing and Big Data*, pp. 56-61, 2018.

[12] S. Parveen, S. N. Deshmukh, "Opinion Mining in Twitter–Sarcasm Detection," *Politics*, vol. 1200, p. 125, 2017.

[13] R. Gupta, J. Kumar, H. Agrawal, "A Statistical Approach for Sarcasm Detection Using Twitter Data," in *Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 633-638, 2020. doi: 10.1109/ICICCS48265.2020.9121043.

[14] Chakravarthi, B. R., N, S., B, B., K, N., Durairaj, T., Ponnusamy, R., Kumaresan, P. K., Ponnusamy, K. K., Rajkumar, C. (2024). Overview of sarcasm identification of Dravidian languages in DravidianCodeMix@FIRE-2024. In Forum of Information Retrieval and Evaluation FIRE - 2024. DAIICT, Gandhinagar.