

CLEF 2024 JOKER Task 2 : Using RoBERTa and Bert-uncased for Humour Classification According to Genre and Technique

Notebook for the Joker Lab at CLEF 2024

Sarvesh Narayanan^{1,†}, Jayasimman J^{1,†} and Shiva Ganesh V^{1,†}

¹SSN College of Engineering

Abstract

Humor classification is a complex task in natural language processing (NLP) that involves identifying and categorizing humor based on its various forms and techniques. This study explores the use of two pre-trained transformer models, RoBERTa and BERT-uncased, for humor classification according to genre and technique. The models are fine-tuned on a dataset annotated with different humor genres (such as sarcasm, irony, exaggerations and witty jokes). The performance of each model is evaluated based on accuracy, precision, recall, and F1-score. Our results demonstrate that both RoBERTa and BERT-uncased are effective in capturing the nuances of humor, with RoBERTa showing a slight edge in overall performance. This research highlights the potential of transformer models in advancing the field of humor classification and provides insights into their applicability for more nuanced and context-aware NLP tasks.

Keywords

Humor Classification, Natural Language Processing (NLP), RoBERTa, BERT-uncased

1. Introduction

Humor is an intricate aspect of human communication that poses significant challenges for natural language processing (NLP). The ability to recognize and classify humor is crucial for enhancing human-computer interactions, particularly in applications like virtual assistants, chatbots, and social media analysis. Despite its importance, automatic humor classification remains a difficult task due to the subjective and context-dependent nature of humor.

1.1. Motivation for Research:

The motivation for this research stems from the need to improve the accuracy and sophistication of humor classification systems. Previous approaches have often struggled to capture the subtle nuances and diverse techniques of humor, which can vary significantly across different genres. By leveraging advanced transformer models such as RoBERTa and BERT-uncased, we aim to develop more robust and nuanced humor classification methods. These models, pre-trained on vast amounts of text data, have shown remarkable performance in various NLP tasks and offer the potential to enhance humor recognition.

1.2. Task Description:

This paper focuses on humor classification according to genre and technique, as outlined in the CLEF 2024 JOKER task 2 [1][2]. The task involves categorizing humorous texts into predefined genres (such as satire, sarcasm, and puns) and techniques (such as wordplay, incongruity, and exaggeration). The

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

[†]These authors contributed equally.

✉ sarveshnarayanan4@gmail.com (S. Narayanan); jayasimmanj27@gmail.com (J. J); shivaviswanathan07@gmail.com (S. G. V)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

JOKER lab, introduced in several key publications, provides a comprehensive dataset for this task, allowing for rigorous evaluation of our models.

1.3. State-of-the-Art Overview:

Recent advancements in NLP have been dominated by transformer-based models, which have achieved state-of-the-art results in numerous tasks. RoBERTa, a robustly optimized BERT approach, has demonstrated superior performance due to its extensive pre-training and fine-tuning capabilities. BERT-uncased, known for its bidirectional training of transformers, excels in understanding context and nuance in text. Prior studies have explored these models for various classification tasks, but their application to humor classification remains under-explored. This study seeks to address this gap by evaluating the performance of RoBERTa and BERT-uncased models in humor classification tasks.

This structured approach aims to provide a comprehensive understanding of the application of transformer models in humor classification, highlighting their strengths and limitations in this nuanced task.

2. Approach

The dataset for this task is provided by the CLEF 2024 JOKER lab, which focuses on automatic humor analysis. The dataset consists of humorous texts categorized according to genre (such as sarcasm, irony, exaggerations and witty jokes). This diverse dataset enables the development and evaluation of models capable of distinguishing between different types and techniques of humor. The two datasets train-input and train-qrels are combined to have the format as in Table 1 .

Table 1

Table with ID, Text, and Class columns.

ID	Text	Class
1661	How do you organize a space party? You planet.	AID

The dataset is pre-processed to ensure consistent formatting, with texts tokenized and annotated for genre and technique. This structured data allows for the effective training and evaluation of machine learning models.

To tackle the humor classification task, we employed two transformer-based models: BERT-uncased and RoBERTa. These models were chosen for their proven capabilities in understanding contextual and semantic nuances in text.

2.0.1. BERT-uncased

BERT [3] (Bidirectional Encoder Representations from Transformers) is a transformer model pre-trained on a large corpus of English text. The uncased version ignores case distinctions, making it robust to variations in capitalization. BERT-uncased is fine-tuned for humor classification by adding a classification layer on top of the pre-trained model.

2.0.2. RoBERTa

RoBERTa [4] (A Robustly Optimized BERT Pretraining Approach) builds on BERT's architecture with improvements in pre-training methodology. RoBERTa undergoes more extensive training on a larger dataset, resulting in enhanced performance.

2.1. Training Setup

- **Data Split:** The dataset is split into training, validation, and test sets with an 80-20 ratio.
- **Batch Size:** 8

- **Learning Rate:** 5e-5
- **Epochs:** 3
- **Optimizer:** Adam
- **Loss Function:** Sparse Categorical Cross-Entropy Loss

2.2. Fine-Tuning Process

The models are fine-tuned using the training set, with hyperparameters optimized based on performance on the validation set. Fine-tuning involves adjusting the weights of the pre-trained model to minimize the classification error.

2.3. Model Evaluation

The performance of the models is evaluated based on their accuracy in distinguishing between different genres and techniques of humor. By evaluating the models based on accuracy, we aim to assess their ability to capture the nuances of humor and classify texts correctly. This metric provides a straightforward measure of the models' performance and their effectiveness in humor classification tasks.

This methodological approach ensures a rigorous and reproducible framework for humor classification, leveraging state-of-the-art NLP models to capture the complexities of humor in text.

3. Results

Our study delved into the realm of humor classification using state-of-the-art transformer models, specifically RoBERTa and BERT-uncased. The results of our experiment showcase the effectiveness of these models in deciphering the intricacies of humor, albeit with some limitations.

Overall, our models demonstrated competitive performance, with RoBERTa slightly surpassing BERT-uncased by achieving an accuracy of 70% compared to 67%. This indicates that both models are adept at grasping the nuances of humor, as evidenced by their ability to classify jokes based on genre and technique with reasonable accuracy.

Table 2
Test Results

model	accuracy	SD_precision	SD_recall	SD_f1-score	SD_support
roberta	0.70	0.37	0.66	0.47	38.00
bert uncased	0.67	0.69	0.84	0.76	91.00

Despite their overall effectiveness, both RoBERTa and BERT-uncased exhibited challenges when confronted with certain types of humor. Exaggeration-style jokes and those heavily reliant on cultural references proved to be particularly problematic for the models. These types of humor often require a deep understanding of context and cultural nuances, which may have posed difficulties for the models' classification abilities.

To address these challenges and further enhance the performance of transformer models in humor classification, several strategies could be explored. One approach could involve incorporating additional contextual information, such as cultural context or background knowledge, into the models. This could potentially help the models better understand and classify jokes that rely heavily on cultural references.

Furthermore, fine-tuning strategies tailored to these specific types of humor could also prove beneficial. By focusing on optimizing the models' performance on exaggeration-style jokes and culturally nuanced humor, we may be able to improve their overall classification accuracy.

As a result, while transformer models like RoBERTa and BERT-uncased show promise for humor classification, there is still room for improvement, especially when dealing with more nuanced forms

of humor. Future research in this area could focus on developing more sophisticated models and fine-tuning strategies to better capture the complexities of humor in text.

4. Discussion

The experiments conducted demonstrate the potential of transformer-based models in humor classification. Both BERT-uncased and RoBERTa showed considerable effectiveness, but several challenges remain. These models struggled with jokes heavily dependent on cultural references or requiring an understanding of specific exaggerations. This indicates a need for incorporating more diverse training data and possibly integrating external knowledge sources to handle such cases better. Additionally, the importance of model interpretability should not be overlooked, as understanding why a model classifies a text as humorous is crucial for further development and trust in AI systems.

5. Future Work

Future work could involve exploring more sophisticated ensemble methods, incorporating additional features such as sentiment and semantic similarity, and applying these models to larger and more diverse datasets. Further investigation into interpretability and the handling of ambiguous or context-dependent humor could also provide valuable insights for enhancing humor classification systems. Enhancing the dataset with more varied and culturally rich humorous texts could also improve the models' robustness and applicability across different contexts and populations.

6. Conclusion

In this paper, we explored the use of advanced transformer-based models, specifically BERT-uncased and RoBERTa, for humor classification according to genre and technique. By leveraging these pre-trained models and fine-tuning them on the provided dataset, we aimed to improve the accuracy and robustness of humor classification systems. Our experiments demonstrated that both models are effective in capturing the nuances of humor, with RoBERTa slightly outperforming BERT-uncased in overall performance metrics.

Key observations from our experiments indicate that transformer models, with their deep contextual understanding and extensive pre-training, can significantly enhance the ability to classify humorous texts. The results also highlight the importance of fine-tuning and hyperparameter optimization in achieving optimal performance.

References

- [1] L. Ermakova, T. Miller, A.-G. Bossler, V. M. P. Preciado, G. Sidorov, A. Jatowt, Overview of joker - clef-2024 track on automatic humor analysis, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, L. Soulier, G. M. D. Nunzio, P. Galuščáková, A. G. S. de Herrera, G. Faggioli, N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024)*, 2024.
- [2] V. M. P. Preciado, et al., Overview of the clef 2024 joker task 2: Humour classification according to genre and technique, in: G. Faggioli, et al. (Eds.), *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024)*, CEUR Workshop Proceedings, CEUR-WS.org, 2024.
- [3] M. Geetha, D. K. Renuka, Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model, *International Journal of Intelligent Networks 2* (2021) 64–69.
- [4] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, *arXiv preprint arXiv:1907.11692* (2019). URL: <https://doi.org/10.48550/arXiv.1907.11692>.