GHaLIB غالب : Generating Hope and Linguistic Irony in Banter

عبداللہ ابراہیم[1] (22L-7552) احمد عبداللہ[1] (22L-7503)

نعمان جعفر[1] (22L-7535)

[1]FAST-NUCES, Department of Data Science & AI

1 Background

Sentiment analysis is an essential field in Natural Language Processing (NLP) that helps understand public sentiment across various domains, including social media, news, and customer feedback. However, analyzing sentiments involving sarcasm and hope speech presents a unique challenge due to the nuanced nature of human language. Sarcasm is particularly difficult to detect because it involves saying something that typically conveys the opposite of its literal meaning. Many sentiment analysis models struggle with sarcasm, often misclassifying sarcastic statements as having a neutral or even positive sentiment when, in reality, they carry a negative undertone.

On the other hand, hope speech represents positive, motivational, and uplifting communication. Detecting hope speech is crucial in social media analysis and mental health applications, as it can help in identifying and amplifying positive discussions. However, hope speech often gets overshadowed by more dominant sentiments, making its detection a non-trivial task. Our project aims to address these challenges by implementing advanced NLP techniques to accurately classify sarcasm and hope speech, contributing to a more comprehensive and fine-grained sentiment analysis system. We will be using the PolyHope at IberLEF 2025 dataset which was released on February 15th, 2025.

2 Literature Review

2.1 *Research on Pattern-Based Sarcasm Detection*

Bouazizi et. al (2016) have been analyzing the detection of sarcasm on Twitter using a pattern-based approach [1]. They collected sarcastic tweets through a query to the Twitter API for tweets containing the hashtag *#sarcasm*, thus forming three balanced datasets with sarcasm being classified into three types:

wit, whimper, and evasion. They used the classification with punctuation-related features, analyzing the contribution of punctuation patterns in sarcasm detection from given texts. Their proposed approach gave an F1 measure of about 82.7% on the Twitter dataset and an accuracy of about 83.46%, thus proving competitive with existing techniques for detecting sarcasm [1]. This study was further extended by Oxana et. al for context features, emotions and texts [2] [3].

2.2 *Research on Low-Resource NLP*

Dave and Desai et. al (2016) has handled a multilingual architecture for sarcasm detection with a focus on Hindi. They worked with the application of SVM classifier with 10-fold cross-validation with the simple Bag-of-Words approach for features and TF-IDF for frequency measurement [4]. The simplistic model classifies only 50% of the sarcastic sentences accurately, implying that the method is extremely insufficient and simple Bag-of-Words is not very effective in classifying sarcasm [3] [5]. They, therefore, added a model with an improved performance lexicon-based feature using both dictionary and corpus techniques to improve the model understanding of sarcasm in Hindi text [4]. Thus, this study is on proving the complexity of detecting sarcasm in a multilingual context while emphasizing the need for more advanced linguistic features than just using simple word frequency as representations [4] [6].

2.3 *Research on Enhancing Sarcasm Detection with Contextual and Emotional Cues*

Vitmana et al. (2023) proposed a sarcasm detection framework, integrating contextual, emotional, and sentiment-based features to enhance classification accuracy. The approach employs transformer-based models along with CNNs to capture linguistic incongruities that characterize sarcasm, thereby improving the training [2]. The framework achieves state-of-the-art performance over a variety of benchmarks, using advanced pre-trained sentiment and emotion analysis models. The findings reveal that emotion detection enhances sarcasm classification; also, combining contextual transformer embeddings with CNNs produces better results than legacy methods. This model performed far better on the Reddit, Internet Argument Corpus, and Twitter datasets, which further facilitates the place of sarcasm detection in focus [2] [6].

2.4 *Research on Sentiment Analysis Using Deep CNNs*

Zhao et al. (2017) tested analyzing tweets for sentiment using deep learning and their methodology tested on five different datasets to find out the performance. A deep Convolutional Neural Network (CNN) was used with a scoring mechanism called *SenScore(w)* in which w refers to a lexicon word [6]. A tweet entered the model and was transformed into an n $\times$ k representation (with n is ***number***

of words and k is ***embedding size***) and was then sent through three convolutional layers and a fully connected (FC) layer in that order [6]. GloVe-DCNN incorporates a regression model that integrates two major paradigms of value in sentiment analysis: one, a local context window, which focuses on words immediately near to it in an utterance, and the other, global matrix factorization, which recognizes a general relationship pattern of words in a huge body of text.

An average operational accuracy of 85.63% was obtained by this approach, beating others by at least 3.69% improvement, signifying that deep CNNs along with efficient word representations can act as boosters for the performance of sentiment analysis [6].

2.5 *Research on ULMFiT and Transfer Learning for NLP Efficiency*

Howard and Ruder et. al (2018) introduced the Universal Language Model Fine-tuning (ULMFiT), a method meant to extend transfer learning in NLP by using pre-trained language models. Conventional NLP models need a lot of labeled data and prove inefficient when trained from scratch, thus making transfer learning a vital element in boosting performance under limited-resource scenarios [7]. ULMFiT addresses these problems by utilizing several fine-tuning strategies including discriminative fine-tuning, where different layers of the model learn at different rates; slanted triangular learning rates, which allow fast convergence of the model; and gradual unfreezing, which has the effect of reducing overfitting by fine-tuning layers in a progressive manner. Their findings showed that pretraining on a general-domain corpus leads to dramatic performance improvements on downstream tasks and helps with overfitting and catastrophic forgetting-the two major challenges confronting transfer learning. ULMFiT was tested on six text classification datasets, where it outperformed the state-of-the-art methods by an error margin of 18-24% improvement [7].

The data-efficiency of the model was impressive, achieving competitive performances with as low as 100 labeled examples which matched the results of models trained on 100 times larger datasets. The study offers insight into the success of transfer learning approaches in the arena of NLP and underlines the role of fine-tuning strategies in improving a model's generalization especially in low-resource settings [7] [6] [8].

2.6 *Research on Multi-Task Learning for Sentiment Analysis*

Tan et. al (2023) introduced a framework for multi-task learning in which a deeply layered BiLSTM neural network jointly performed sentiment analysis and sarcasm detection [9]. By simultaneously learning sentimental polarity and cues for sarcasm, the model minimizes mis-classification errors due to sarcasm. The results indicate that multi-task learning promotes efficiency, improving the score of the BiLSTM model by 3% as compared to techniques for classical sentiment analysis. Therefore, this approach has a state-of-the-art boost in sarcastic sentiment classification accuracy, resulting in even more stronger and interpretable sentiment analysis [9].

2.7 *Research on Sentiment Analysis Using LLMs*

Zhang and Deng et. al (2024) have studied how Large Language Models (LLMs) fare in sentiment analysis compared to Small Language Models (SLMs) that are trained on domain-specific datasets. They compared the models on 13 different tasks associated with 26 datasets for a final evaluation by sampling up to 500 examples from each test set for balanced evaluations. For the LLM part, they employed two Flan family models, Flan-T5 (XXL version, 13B) and Flan-UL2 (20B) [8], for the very reason that they have excellent zero and few-shot properties for their testing [10]. Their results state that LLMs, along with models like ChatGPT [11], perform excellently well in zero-shot settings and can handle sentiment-analysis tasks without prior fine-tuning on specific datasets [12].

Most importantly, LLMs, unlike SLMs, were very effective in few-shot learning contexts, which implies their incredible potential in cases of limited labeled training data. This shows that LLMs are highly promising for real-world sentiment analysis applications, which involve expensive and few labeled data [10].

2.8 *Research on Zero-Shot Classification*

In Puri's and Catanzaro's (2019) investigation of generative language models for zero-shot text classification, natural language classifications were provided instead of task-oriented training examples. Such fine-tuning on title prediction tasks allowed GPT-2 [11] to generalize across benchmarks, and it was shown that zero-shot learning outperformed the random and majority class baselines. By reframing classification problems into multiple-choice ones, adaptability was improved; pretraining on an extensive scale using datasets endowed with metadata ensured good generalization [13].

2.9 *Research on Prompt-based Classification*

In 2023, Sun et al. proposed a framework, Clue And Reasoning Prompting (CARP), giving the clue identification, diagnosis reasoning, and decision-making process in order to enhance LLM-based text classification. Compared to existing methods of in-context learning, CARP produced superior results on elaborate linguistic tasks in low-resource settings while requiring far fewer labeled examples than traditional models [14].

2.10 *Research on Implicit Sentiment Analysis*

Xu et. al (2022) proposed KC-ISA, an implicit sentiment analysis model that integrates both knowledge enhancement and contextual features to assist in sentiment classification. Unlike sentiment analysis, which deals with the direct name of sentiment, implicit sentiment analysis requires a deep understanding of linguistic lingo and contextual understanding to make inferences [15][4]. To counter this challenge, the researchers designed a Knowledge Fusion Module which integrates external knowledge from XLORE. Using this external knowledge, the system is enabled to infer implicit sentiment expression in the absence of direct emotional clues [15] [3].

The KC-ISA model was evaluated on the SMP2019 implicit sentiment analysis dataset whereby these authors uncovered three key dependencies representing potential sources of improvement for implicit detection. The first, context dependency, means that a text surrounding a given sentiment would determine whether the sentiment is positive or negative; whilst sentiment target dependency keeps track of whether a sentiment has been assigned to the correct target; finally, knowledge and common sense dependency exploit knowledge concerning such external objects which could help in qualifying the sentiment [12]. The integration of dependency factors is what made the KC-ISA model highly efficiently, and it achieved an accuracy of 0.786 and F1 of 0.755 for implicit sentiment classification [15].

2.11 *Research on RGPT for LLM-Based Classification*

Zhang et. al (2024) proposed RGPT, which works similar to an adaptive boosting framework for text classification-the LLMs are fine-tuned iteratively while at the same time the distributions of the training samples are modified. The RGPT builds a highly specialized classifier by the ensemble of several fine-tuned models [16]. It goes beyond the performance of eight strong contenders PLM and seven versions of LLMs across four benchmarks and above-average human classification performance. The adaptive boosting technique generalizes better by focusing misclassified samples and therefore significantly enhances LLM-based text classification [16].

References

[1] M. Bouazizi and T. Otsuki Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016. DOI: 10.1109/ACCESS.2016.2594194.

[2] O. Vitman, Y. Kostiuk, G. Sidorov, and A. Gelbukh, *Sarcasm detection framework using context, emotion and sentiment features*, 2023. arXiv: 2211.13014 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2211.13014.

[3] P. Verma, N. Shukla, and A. Shukla, "Techniques of sarcasm detection: A review," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021, pp. 968–972. DOI: 10.1109/ICACITE51222.2021.9404585.

[4] A. D. Dave and N. P. Desai, "A comprehensive study of classification techniques for sarcasm detection on textual data," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 1985–1991. DOI: 10.1109/ICEEOT.2016.7755036.

[5] Y. Du, T. Li, M. S. Pathan, H. K. Teklehaimanot, and Z. Yang, "An effective sarcasm detection approach based on sentimental context and individual expression habits," *Cognitive Computation*, vol. 14, no. 1, pp. 78–90, 2022, ISSN: 1866-9964. DOI: 10.1007/s12559-021-09832-x. [Online]. Available: https://doi.org/10.1007/s12559-021-09832-x.

[6] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis," *IEEE Access*, vol. 6, pp. 23 253–23 260, 2018. DOI: 10.1109/ACCESS.2017.2776930.

[7] J. Howard and S. Ruder, *Universal language model fine-tuning for text classification*, 2018. arXiv: 1801.06146 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1801.06146.

[8] H. W. Chung, L. Hou, S. Longpre, *et al.*, *Scaling instruction-finetuned language models*, 2022. arXiv: 2210.11416 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2210.11416.

[9] Y. Y. Tan, C.-O. Chow, J. Kanesan, J. H. Chuah, and Y. Lim, "Sentiment analysis and sarcasm detection using deep multi-task learning," *Wireless Personal Communications*, vol. 129, no. 3, pp. 2213–2237, 2023, ISSN: 1572-834X. DOI: 10.1007/s11277-023-10235-4. [Online]. Available: https://doi.org/10.1007/s11277-023-10235-4.

[10] W. Zhang, Y. Deng, B. Liu, S. Pan, and L. Bing, "Sentiment analysis in the era of large language models: A reality check," in *Findings of the Association for Computational Linguistics: NAACL 2024*, K. Duh, H. Gomez, and S. Bethard, Eds., Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 3881–3906. DOI: 10.18653/v1/2024.findings-naacl.246. [Online]. Available: https://aclanthology.org/2024.findings-naacl.246/.

[11] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention is all you need*, 2023. arXiv: 1706.03762 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1706.03762.

[12] A. S. Talaat, "Sentiment analysis classification system using hybrid bert models," *Journal of Big Data*, vol. 10, no. 1, p. 110, 2023, ISSN: 2196-1115. DOI: 10.1186/s40537-023-00781-w. [Online]. Available: https://doi.org/10.1186/s40537-023-00781-w.

[13] R. Puri and B. Catanzaro, *Zero-shot text classification with generative language models*, 2019. arXiv: 1912.10165 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1912.10165.

[14] X. Sun, X. Li, J. Li, *et al.*, *Text classification via large language models*, 2023. arXiv: 2305.08377 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2305.08377.

[15] M. Xu, D. Wang, S. Feng, Z. Yang, and Y. Zhang, "KC-ISA: An implicit sentiment analysis model combining knowledge enhancement and context features," in *Proceedings of the 29th International Conference on Computational Linguistics*, N. Calzolari, C.-R. Huang, H. Kim, *et al.*, Eds., Gyeongju, Republic of Korea: International Committee on Computational Linguistics, Oct. 2022, pp. 6906–6915.

[16] Y. Zhang, M. Wang, C. Ren, *et al.*, *Pushing the limit of llm capacity for text classification*, 2024. arXiv: `2402.07470 [cs.CL]`. [Online]. Available: `https://arxiv.org/abs/2402.07470`.