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 Research Problem and Motivation

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.

2 NLP Techniques

To effectively detect sarcasm and hope speech, we will leverage several modern NLP techniques.

- **Large Language Models (LLMs):** Pre-trained LLMs such as BERT, RoBERTa, or GPT-based models will be fine-tuned on our datasets to improve their ability to distinguish between literal and non-literal meanings in text. Fine-tuning these models will help adapt them to the specific linguistic patterns associated with sarcasm and hope speech.

- **Transformers:** Since sarcasm and hope speech often require an understanding of context, we will use transformer-based models that capture long-range dependencies between words in a sentence. These models are highly effective at handling complex linguistic structures, improving classification accuracy.
- **Retrieval-Augmented Generation (RAG):** By integrating retrieval-based methods with generative capabilities, we aim to enhance model performance by allowing it to pull relevant contextual information before making predictions. This is particularly useful for sarcasm detection, where external knowledge can sometimes be necessary to understand the intended meaning.

3 Datasets and Tools

To train and evaluate our models, we will utilize three key datasets:

1. **Codabench Competition Dataset (Primary)** – Our ultimate objective is to transition to this dataset, which is sourced from the PolyHope competition at IberLEF 2025: *Optimism, Expectation, or Sarcasm?* This dataset is specifically designed for hope speech sentiment analysis and will allow us to build models capable of distinguishing between various forms of hopeful expression, as well as detecting sarcasm that mimics optimism.

   The dataset is structured into two primary subtasks:

   - **Subtask 1: Binary Hope Speech Detection** (English and Spanish) – This task involves identifying both explicit and subtle expressions of hope, even when masked by complex language structures or implied sentiments. The goal is to classify social media texts into two categories:
     - **Hope** – Tweets that convey an expression of hope, expectation, or desire.
     - **Not Hope** – Tweets that do not express hope, expectation, or desire.
   - **Subtask 2: Multiclass Hope Speech Detection** (English and Spanish) – This subtask aims to differentiate between distinct types of hope while also identifying sarcastic expressions that resemble hopeful language. The classification categories include:
     - **Generalized Hope** – Broad, non-specific expressions of hopefulness or optimism not directed at specific events or outcomes.
     - **Realistic Hope** – Expressions of hope based on plausible and meaningful expectations.
     - **Unrealistic Hope** – Expressions of hope for outcomes that are highly unlikely or impossible.

- **Not Hope** – Texts devoid of hopeful sentiment.
- **Sarcasm** – Texts that outwardly express hope but are actually sarcastic in nature.

By leveraging this dataset, we aim to refine our model's ability to differentiate between genuine hope, misplaced optimism, and sarcasm, thereby improving sentiment classification accuracy in social media discourse.

2. **News Headlines Dataset for Sarcasm Detection** – This dataset contains news headlines labeled as sarcastic or non-sarcastic. It will help us understand how sarcasm manifests in formal news contexts.

3. **Tweets with Sarcasm and Irony** – This dataset consists of tweets labeled for sarcasm and irony, providing insights into how sarcastic language is used in informal, real-world conversations on social media.

We plan to use Python (3.12 onwards) for coding, utilizing libraries PyTorch, ONNX, TensorFlow/Keras, and HuggingFace Datasets to facilitate model training and evaluation. We will be using Kaggle GPUs (2*T4 Tesla and 4*L4 Tesla) and Kaggle Notebooks.

4 Evaluation Metrics

To ensure that our models are performing optimally, we will use a combination of evaluation metrics:

To ensure that our models are performing optimally, we will use a combination of evaluation metrics:

1. **Accuracy and pAUC(80*%)**: Measures the overall correctness of our model's predictions. It will be considered as our primary metric for benchmark and paper.

2. **Precision, Recall, and F1-Score; Combined Metric**: These metrics will be essential, as they provide a more detailed analysis of the model's ability to distinguish between sarcastic, non-sarcastic, and hopeful speech.

$$OverallScore = 0.3 \times Precision + 0.6 \times F1Score + 0.1 \times Recall \quad (1)$$

3. **Confusion Matrix**: This will help us analyze where the model is making errors, such as misclassifying sarcasm as a neutral sentiment.

References

[1] R. Misra, "News Headlines Dataset for Sarcasm Detection," 2019. [Online]. Available: https://www.kaggle.com/datasets/rmisra/news-headlines-dataset-for-sarcasm-detection.

[2] N. Johnk, ”Tweets with Sarcasm and Irony,” 2019. [Online]. Available: https://www.kaggle.com/datasets/nikhiljohnk/tweets-with-sarcasm-and-irony.

[3] ”Codabench Competition on Hope Speech Sentiment Analysis.” [Online]. Available: https://www.codabench.org/competitions/5509/.

[4] A. Joshi, P. Bhattacharyya, and M. J. Carman, ”Automatic Sarcasm Detection: A Survey,” arXiv preprint arXiv:1602.03426, 2016. [Online]. Available: https://arxiv.org/abs/1602.03426.

[5] R. Badlani, N. Asnani, and M. Rai, ”An Ensemble of Humour, Sarcasm, and Hate Speech for Sentiment Classification in Online Reviews,” in *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, Association for Computational Linguistics, pp. 337–345, 2019. [Online]. Available: https://aclanthology.org/D19-5544/.