

Context-Aware Sarcasm and Irony Detection in English, Hinglish and optionally Hindi

Aanvik Bhatnagar (2022101103)
aanvik.bhatnagar@students.iiit.ac.in

Harshit Gupta (2022101124) Manan Garg (2022101109)
harshit.gupta@students.iiit.ac.in manan.garg@students.iiit.ac.in

Team: Unknown_Tokens
ANLP M25, IIIT Hyderabad

Abstract

Detecting sarcasm and irony in online text is a significant challenge for NLP, particularly in code-mixed languages like Hinglish. This paper introduces a dual-track framework to address this. For English, we employ a RoBERTa model augmented with a Recurrent Convolutional Neural Network (RCNN) to capture contextual cues. For Hinglish, we use a multilingual BERT-GRU architecture enriched with semantic knowledge from ConceptNet to handle cultural and linguistic nuances. We leverage Low-Rank Adaptation (LoRA) for efficient fine-tuning and propose an ensemble method to integrate both tracks. Our framework will be evaluated on benchmark datasets using F1-score, precision, and recall to demonstrate the effectiveness of our context-aware, multilingual approach. To address data scarcity, we also optionally plan to introduce a novel data augmentation pipeline to mine sarcastic sentences from large monolingual Hindi corpora using translation and semantic similarity search.

1 Problem Statement and Motivation

Sarcasm and irony are pervasive in online discourse, particularly on platforms such as Twitter, Reddit, and multilingual social media in India. These figurative forms of language often reverse the intended sentiment, making standard sentiment analysis models fail. For example, “*Great, another power cut in summer*” appears lexically positive but conveys frustration. Such cases are not rare - they constitute a significant portion of online conversations and user-generated reviews.

The challenge is twofold:

1. **Context-dependence:** Non-literal meaning depends on pragmatic and discourse-level cues.

2. **Multilingual setting:** In Hinglish (code-mixed Hindi-English), sarcasm uses cultural references, transliterated tokens, and informal spellings that break standard NLP pipelines.

This project aims to build a **flexible, context-aware framework** for sarcasm and irony detection in English, Hinglish and optionally Hindi using transformer-based methods, augmented with contextual, semantic, and ensemble strategies.

2 Background and Related Work

Several recent studies have explored transformer-based approaches to sarcasm detection:

1. **RCNN + RoBERTa:** Combining RoBERTa embeddings with recurrent convolutional layers captures temporal dependencies missed by vanilla transformers. This hybrid model achieved state-of-the-art results across multiple sarcasm datasets.
2. **SemIRNet (Semantic Irony Recognition Network):** Incorporates ConceptNet knowledge graphs for multimodal sarcasm detection. Semantic enrichment improves performance beyond feature alignment, showing the role of external commonsense.
3. **BERT-GRU for Hinglish:** Leveraging multilingual BERT with a GRU layer, this approach effectively captures sequential dependencies in noisy Hinglish text, achieving strong F1 scores (>0.95).
4. **Data Augmentation via Translation and Embedding:** Techniques like back-translation are common for data augmentation. More advanced methods leverage sentence embeddings (e.g., from LaBSE or Laser3) to find paraphrases or semantically similar sentences in large unlabeled corpora.

These findings suggest that while pretrained transformers are strong baselines, additional context modeling (RCNN/GRU) and knowledge augmentation (ConceptNet) further improve robustness - particularly crucial for Hinglish.

3 Proposed Objectives

The primary objectives of this project are outlined as follows:

- Develop a **dual-track system** for English and Hinglish sarcasm/irony detection.
- Investigate **RCNN-RoBERTa**, **BERT-GRU**, and **SemIRNet-style knowledge augmentation**.
- Explore **LoRA (Low-Rank Adaptation)** to efficiently fine-tune large pretrained models.
- Design an **ensemble framework** that integrates English and Hinglish outputs for mixed-language inputs.
- **(Optional)** Explore a novel data augmentation technique using translation and semantic similarity search to generate a new Hindi sarcasm dataset from large monolingual corpora.
- Evaluate models comprehensively across benchmark datasets using F1, precision, recall, and AUC.

4 Dataset

We will experiment with both English and Hinglish sarcasm/irony datasets

English:

1. *Tweets with Sarcasm and Irony*: [Kaggle](#)
2. *Dataset from isarcasm competition*: [Kaggle](#)
3. *SemEval-2018 Task 3*: Irony detection in English tweets Dataset

Hinglish:

1. *HackArena Multilingual Sarcasm Detection*: [Kaggle](#)
2. *Sarcasm Detection Code Mixed Dataset*: [GitHub](#)
3. *Sarcasm Detection in Hindi-English Code-Mixed Data*: [GitHub](#)

Hindi (Optional): To support our data augmentation strategy, we will utilize large monolingual Hindi corpora sourced from:

1. IndicCorp2
2. The Hindi portion of the BPCC (English-Hindi Parallel Text)
3. The Oscar Dataset (Hindi subset)

These will be used to mine for naturally occurring sarcastic sentences.

5 Proposed Methodology

5.1 English Track

- **Base model:** RoBERTa
- **Architecture:** RoBERTa \rightarrow RCNN (captures contextual dependencies)
- **Variants:** RoBERTa + BERT, RoBERTa + XLNet, RoBERTa + LLaMA (comparative baselines)

5.2 Hinglish Track

- **Base models:** IndicBERT and multilingual BERT
- **Architecture:** BERT \rightarrow GRU (captures sequence dependencies in noisy code-mixed input)
- **Semantic Augmentation:** SemIRNet-style enrichment using ConceptNet concepts for cultural markers.
- **(Optional)** Extension to Hindi Dataset

5.3 Hinglish-English Ensemble

- Weighted averaging of predictions from English + Hinglish systems.
- Optionally experiment with **stacked ensemble** (meta-classifier over both tracks).

5.4 Efficient Fine-tuning

- Apply **LoRA adapters** on all transformer backbones to enable faster experimentation and resource-efficient training.

5.5 (Optional) Data Augmentation via Semantic Search for Hindi

Recognizing the scarcity of labeled Hindi sarcasm data, we propose a novel data augmentation pipeline to create a new dataset. This process involves the following steps (suggested by the mentor):

1. **Translation:** We will take an existing English sarcasm and irony dataset (denoted as Sar_E) and translate it into Hindi using a high-quality machine translation model. The resulting dataset is Sar_H' .
2. **Monolingual Corpus Aggregation:** We will compile a large monolingual Hindi corpus ($Mono_H$) by combining text from IndicCorp2, BPCC and the Oscar dataset.

3. **Semantic Similarity Search:** Using a powerful sentence embedding model such as Laser3 or LaBSE, we will generate vector representations for all sentences in both Sar_H' and $Mono_H$. We will then perform a similarity search to find sentences in $Mono_H$ that are semantically close to the translated sarcastic sentences in Sar_H' .
4. **Dataset Creation:** The set of semantically similar sentences retrieved from the monolingual corpus will form our new, potential Hindi sarcasm dataset, $Mono_H'$. The remaining sentences ($Mono_H - Mono_H'$) can serve as a large non-sarcastic dataset.
5. **Manual Verification:** A crucial final step involves manually sampling and reviewing sentences from $Mono_H'$ to validate whether they are genuinely sarcastic. This quality check will determine the viability of using this augmented dataset for training the Hinglish track models.

6 Evaluation

- Metrics: Precision, Recall, F1-score (macro and class-wise), ROC-AUC.
- Baselines: Compare against plain RoBERTa, plain BERT, XLNet, and multilingual BERT.
- Ablation studies:
 - Effect of RCNN/GRU layers.
 - Effect of knowledge augmentation (ConceptNet).
 - Effect of ensemble averaging.
- Cross-lingual testing: Apply English model on Hinglish data (and vice versa) to measure transfer.

7 Tentative Timeline

We propose the 9-week timeline for the project mentioned in Table 1. Based on the progress of these deliverables, we will alter the timeline to incorporate the optional deliverable of data augmentation for Hindi and then apply sarcasm detection to this dataset with another model.

Week	Task
1	Lit. Review & Dataset Finalization & cleaning
2	Baseline implementations (RoBERTa, BERT)
3	RCNN-RoBERTa setup & initial English experiments
4	Hinglish track: BERT-GRU baseline
5	Incorporation of ConceptNet for semantic augmentation
6	Apply LoRA adapters for efficient fine-tuning
7	Ensemble integration (weighted + meta-classifier)
8	Full evaluation (ablation studies, cross-lingual tests)
9	Report writing: interim + final proposal

Table 1: Project Timeline.

- [2] Liang Jing, Jing Li, Zheng-Jun Zha, Meng-Hao Guo, Yuchao Feng, and Qing-Guo Chen. 2022. [SemIRNet: A Semantic Irony Recognition Network for Multimodal Sarcasm Detection](#). In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*.
- [3] Anmol Gupta and Ramalingam Ponnusamy. 2021. [Hilarious or Hidden? Detecting Sarcasm in Hinglish Tweets using BERT-GRU](#). In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*.

References

- [1] Michail Potamias, Georgios Siolas, and Andreas-Georgios Stafylopatis. 2020. [A transformer-based approach to irony and sarcasm detection](#). In *Proceedings of the 11th Hellenic Conference on Artificial Intelligence (SETN 2020)*.