

Advanced NLP M25

Project Proposal: Unknown Tokens

Title: A Flexible Framework for Context-Aware Sarcasm and Irony Detection in English and Hindi

Project Area(s)

- Interpretability/Explainability
- NLP Applications
- Document Understanding

Problem Statement

Sarcasm and irony are pervasive in online communication, posing a significant challenge for automated text understanding. Their meaning is highly **context-dependent** and often expressed through contradictions, making them difficult to detect. Standard sentiment analysis models frequently fail because they interpret text literally, misclassifying positive words used to convey a negative sarcastic intent. For instance, the phrase "Great, another meeting" is positive on the surface but often expresses frustration.

Effectively identifying such non-literal language requires moving beyond simple lexical cues to a deeper **pragmatic interpretation**. NLP methods are uniquely suited for this task as they can model the complex interplay between words, conversational context, and discourse structure. This challenge is further amplified in a multilingual setting, as the linguistic and cultural markers of sarcasm can differ significantly between languages like **English and Hindi**.

This project aims to develop a flexible and robust framework for sarcasm and irony detection that leverages linguistic features and contextual information. The framework will be designed to operate effectively in both English and Hindi, addressing the unique challenges of non-literal language understanding in each language independently.

Proposed Objectives

- **Develop an English Sarcasm Model:** Build a robust model to accurately detect sarcasm and irony in English text.
- **Integrate Contextual Analysis:** Improve model performance by analyzing conversational context and discourse-level features beyond single sentences.
- **Build an Independent Hindi Model:** Create a sarcasm detection model specifically for Hindi from scratch, focusing on language-specific data curation and feature engineering.
- **Explore Implicit Knowledge (Optional):** As an exploratory goal, investigate how "common sense" or world knowledge can be used to improve detection accuracy.
- **Design a Modular Framework:** Create a flexible architecture that can be easily extended to new languages, domains, or data types in the future.