Advancing Sarcasm Detection with Metric Learning and Contextual Embeddings

Bala Harini Ramesh, Avyay Kodali

Abstract

This paper explores the use of an enhanced version of BERT (Bidirectional Encoder Representations from Transformers) and metric space learning in the task of sarcasm detection. BERT embeds text into high-dimensional vector spaces by tokenizing input text and applying a transformer-based architecture with self-attention mechanisms. This enables the model to capture bidirectional contextual relationships between words, resulting in rich, context-aware embeddings. Metric space learning is then employed to refine the model by learning sarcasm-specific distances between sentences. Sarcastic expressions are clustered closer together, while non-sarcastic ones are placed farther apart, improving the model's ability to detect sarcasm across diverse contexts and vocabularies.

1 Introduction

Sarcasm detection has become an important task in natural language processing (NLP) due to the complexity of interpreting sarcastic remarks, which often involve contrasts between literal meanings and intended messages. This subtle form of communication is prevalent in various contexts such as social media, conversation threads, and customer reviews. Traditional sentiment analysis models often fail to recognize sarcasm because they interpret sentences based on their face value, without accounting for the underlying irony or mockery.

Metric learning presents a robust framework to address these challenges by learning the relationships between sarcastic and non-sarcastic text. The model is designed to map text representations into a high-dimensional space where sentences with similar sarcastic tones are positioned closer to each other, while non-sarcastic sentences are placed farther apart. This is done by transforming sentences into vector representations using state-of-the-art embeddings like BERT (Bidirectional Encoder Representations from Transformers)[1][2], Word2Vec[3], or GloVe[4][5]. These embeddings capture the semantic relationships between words, phrases, and sentences, allowing the system to compute distances in the learned space.

BERT (Bidirectional Encoder Representations from Transformers) revolutionized Natural Language Processing (NLP) by introducing a bidirectional model that considers the entire context of a word in a sentence by looking