Fine-Tuned Roberta-Moid Model for Emotion Classification

Dhanesh Kumar & Danish Nanjiani

May 7, 2025

Abstract

In recent years, Natural Language Processing (NLP) has become increasingly popular, enabling machines to comprehend and process human language. One challenging aspect of NLP has been emotion detection, which saw significant advancements with the advent of deep learning models. While Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were initially utilized to capture positional information within text, they proved computationally intensive and slow. The introduction of transformers, such as BERT and RoBERTa, revolutionized NLP by effectively handling positional encodings and attention mechanisms. However, existing models faced challenges in accurately analyzing emotions in movie datasets like xed-en-fi. To address this, we propose the Roberta-Moid model, which leverages transfer learning, architectural enhancements, and optimization algorithms like AdamW to improve sentiment analysis efficiency.

1 Introduction

In recent years, natural language processing (NLP) has witnessed significant advancements, driven largely by the development of powerful transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers). Building upon the success of BERT, researchers at Facebook AI introduced RoBERTa (Robustly optimized BERT approach) as an extension that further improves upon the original architecture.

1.1 RoBERTa Model

RoBERTa retains the core architecture of BERT while introducing key modifications in its pre-training procedure. Unlike BERT, which was pre-trained using a masked language modeling (MLM) objective, RoBERTa is trained on a much larger corpus of text data with dynamically changing masking patterns. This approach helps RoBERTa learn more effectively from unlabeled text and achieve better generalization performance across various downstream NLP tasks.

1.2 Architecture and Pre-Training

The architecture of RoBERTa consists of multiple transformer layers, each incorporating self-attention mechanisms to capture contextual information from input text. During pre-training, RoBERTa is exposed to a diverse range of text data, allowing it to learn robust representations of language semantics and syntax. The pre-training process involves tasks such as masked language modeling, next sentence prediction, and document-level prediction, which collectively enhance the model's ability to understand and generate coherent text.

1.3 Use Case: Emotion Classification in Movie Datasets

One compelling use case for RoBERTa and similar NLP models is emotion classification in movie datasets. Understanding the emotional content of movie dialogues, reviews, and scripts is crucial for various applications, including content recommendation, audience engagement analysis, and sentiment analysis in movie reviews. By fine-tuning RoBERTa on movie-specific emotion datasets, we can develop models capable of accurately identifying and analyzing the emotions expressed in textual data related to movies.

1.4 Significance

The development of RoBERTa represents a significant milestone in the field of NLP, as it demonstrates the effectiveness of large-scale pre-training and transfer learning for improving model performance on a wide range of tasks. By leveraging vast amounts of unlabeled text data and optimizing pre-training procedures, RoBERTa achieves state-of-the-art results on benchmark datasets and continues to push the boundaries of what is possible in NLP research.

1.5 Motivation for Fine-Tuning

While RoBERTa exhibits impressive performance on many NLP tasks, its generic pre-trained representation may not always capture task-specific nuances effectively. Fine-tuning RoBERTa on domain-specific datasets enables further customization and adaptation to specific applications, such as emotion classification in movie datasets. In this work, we introduce the RobertaMoid framework, a fine-tuned version of RoBERTa tailored for emotion recognition in movie-related text.

2 Architecture

The RobertaMoid framework builds upon the RoBERTa architecture with customized modifications tailored for the emotion classification task. The architecture comprises several key components:

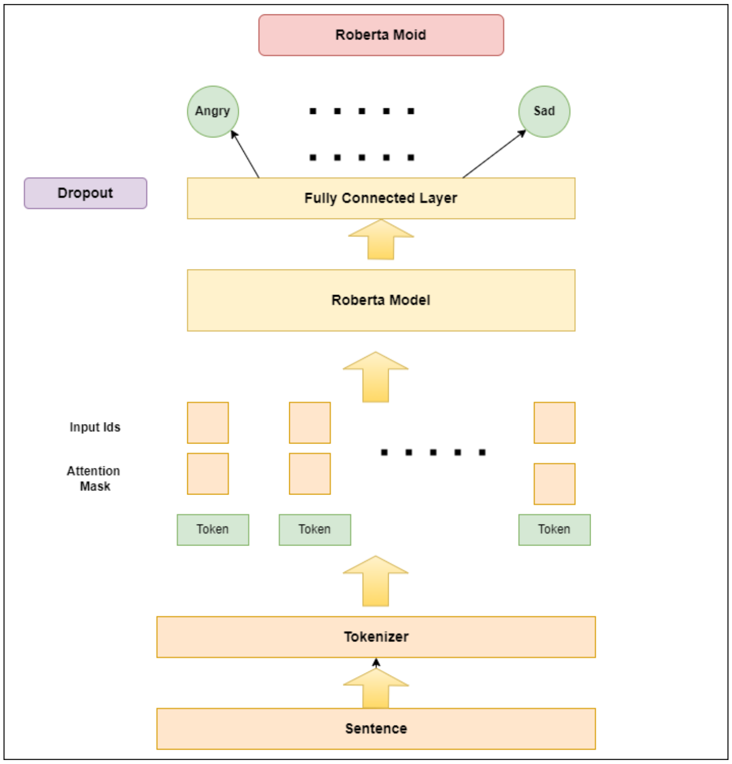

Figure 1: Architecture of the Model.

2.1 RoBERTa Base

The RoBERTa base architecture serves as the foundation of our framework. It consists of multiple transformer layers that capture rich contextual information from input text. These transformer layers employ self-attention mechanisms to effectively encode the semantics and relationships within the text.

2.2 Fully Connected Layer

To adapt RoBERTa for emotion classification, we introduce a fully connected layer after the last transformer layer. This additional layer acts as a feature extractor, transforming the contextualized representations learned by RoBERTa into a format suitable for emotion prediction. By incorporating a fully connected layer, we enable the framework to learn task-specific features relevant to emotion recognition.

2.3 Dropout Layer

To prevent overfitting and improve the generalization capability of the model, we insert a dropout layer after the fully connected layer. The dropout layer randomly drops a certain proportion of the neurons' outputs during training, forcing the network to learn redundant representations and reducing the risk of overfitting to the training data.

2.4 Classification Layer

Following the fully connected layer, we append a classification layer responsible for mapping the extracted features to the output space. The classification layer consists of eight nodes, each corresponding to one of the eight emotion classes present in the dataset. We employ sigmoid activation to obtain probability distributions over the emotion classes, facilitating multi-label classification.

3 Training

We train the RobertaMoid framework using the Xed-en-fi dataset, a curated collection of movie-related text samples annotated with their corresponding emotions. During the training process, we optimize the framework parameters to minimize the binary cross-entropy loss between the predicted probabilities and the ground truth labels. We employ standard techniques such as AdamW or Adam optimization provided by pytorch with a suitable learning rate schedule to efficiently train the framework.

4 Results

The fine-tuned RobertaMoid framework demonstrates notable improvements in emotion classification performance compared to the pre-trained RoBERTa model. We evaluate the framework's performance using various standard metrics including accuracy, precision, and recall. Additionally, we conduct qualitative analysis and compare the framework's predictions with human-labeled ground truth to assess its efficacy in capturing nuanced emotional expressions in movie-related text.

5 Conclusion

The RobertaMoid framework represents a significant advancement in emotion classification for movie datasets. By fine-tuning the RoBERTa architecture and incorporating task-specific modifications, we achieve enhanced accuracy and robustness in identifying emotions expressed in textual data. The framework's versatility and effectiveness make it a valuable tool for applications requiring emotion recognition, ranging from sentiment analysis in movie reviews to personalized content recommendation systems.