Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory

Problem Statement

The exponential growth of user-generated content on platforms like IMDb poses challenges in understanding public opinion at scale. Traditional sentiment analysis techniques often fail to capture long-range dependencies in text, leading to reduced accuracy in sentiment classification. There is a pressing need for models that can efficiently analyze sequential data while addressing issues such as contextual ambiguity and the vanishing gradient problem.

Objectives

This project aims to develop a sentiment analysis system using Long Short-Term Memory (LSTM) networks to classify IMDb movie reviews as positive or negative. The objectives are:
1. To preprocess and clean raw textual data for effective model training.
2. To implement an LSTM-based deep learning model for sentiment classification.
3. To evaluate the model's performance in terms of accuracy and reliability.

Proposed Methodology/Approach

The dataset, consisting of 50,000 IMDb reviews equally divided into positive and negative categories, will be preprocessed through tokenization, stop-word removal, and stemming. Reviews will be vectorized using Doc2Vec for feature extraction. An LSTM classifier optimized with Adam will be trained on 80% of the dataset and validated using 10-fold cross-validation. Performance will be measured using a confusion matrix and accuracy metrics.

Expected Outcome

The project is expected to achieve an accuracy close to 89–90%, demonstrating the potential of LSTM networks for sentiment analysis. The model can be extended to other domains, providing valuable insights for businesses, media platforms, and government applications.

Base Research Paper Reference

Qaisar, S. M. (2020). *Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory*. IEEE ICCIS. Link