Sentiment analysis through CNN.

Team: 33
Members:
MD. Arafat Muktadir 20101373
Md.Mahir Ahnaf Ahmed 20301083
RA:EHSANUR RAHMAN RHYTHM

Introduction

- Introducing a cutting-edge Natural Language Processing (NLP) text classification approach using deep learning models, specifically leveraging a Convolutional Neural Network (CNN).
- The model aims to categorize sentences as simple, complex, or compound, providing a comprehensive understanding of linguistic nuances.
- Our approach combines the power of traditional machine learning algorithms with state-of-the-art neural networks, integrating syntactic and semantic information.

Research Objectives

- Using CNN, the popular deep learning technique in NLP to understand how if performs in NLP tasks.
- Find for better accuracy of classifying sentiment from texts.

Literature Review

- "EDA on Text Classification Tasks" introduces a practical solution to the challenge of limited training data in NLP tasks by proposing four straightforward text editing operations which significantly enhances classification performance across various benchmark tasks.
- Proposed a unified neural network architecture and learning algorithm for various natural language processing tasks, emphasizing the avoidance of task-specific engineered features and the reliance on internal representations learned from vast amounts of data.

Dataset

The IMDb Movie Reviews dataset is a binary sentiment analysis dataset consisting of 50,000 reviews from the Internet Movie Database (IMDb) labelled as positive or negative. The dataset contains an even number of positive and negative reviews. Only highly polarizing reviews are considered. A negative review has a score ≤ 4 out of 10, and a positive review has a score ≥ 7 out of 10. No more than 30 reviews are included per movie. The dataset contains additional unlabelled data.

Methodology

- Here, our proposed methodology consists of six parts. First, we will collect the dataset. Then, we will do preprocessing on the dataset. To elaborate, we will use embedding and set the word limit to a constant for every sentence. Also, we will split the dataset into 2 parts. They are Training and Testing. The ratio would be 5:5. Now, we will be training the deep learning model. After that, we will be testing the deep learning model. Finally, we will acquire the results and analyse them.

Conclusion

We have used a popular deep learning technique called CNN to see how it performs in text classification tasks. We look forward to improving this accuracy and make the process more efficient.