

SENTIMENT ANALYSIS ABOUT IMDB

MOVIE REVIEW

Sunny Dao

BACKGROUND

A. MULTI-LAYER PERCEPTRON (MCP)

1943: Introduced by Warren McCulloch & Walter Pitts.

1957:Frank Rosenblatt developed the “Perceptron”

1960s-1980s: Research in neural networks faced limitations

1986: David Rumelhart, Geoffrey Hinton, and Ronald Williams, enabling training of deep neural networks

Late 1990s-2000s: The term "Multi-Layer Perceptron" (MLP) emerged

2000s-2010s: Advances in hardware, algorithms, and data availability rekindled interest in deep learning, with MLP as a foundational model.

B. CONVOLUTIONAL NEURAL NETWORK(CNN)

1962: Hubel and Wiesel's studies on cat visual cortex inspired the concept of receptive fields and hierarchical visual processing.

1980s: Yann LeCun's work on convolutional neural networks began

1998: LeCun's LeNet-5 demonstrated handwritten digit recognition using convolutional layers.

2012: Alex Krizhevsky's "AlexNet" won the ImageNet competition

2010s: CNNs became a cornerstone of computer vision, achieving state-of-the-art results in various image-related tasks.

2018: CNNs were adapted for other domains like natural language processing (NLP) with models like BERT

INTRODUCTION

In the realm of modern entertainment, movies hold a significant place in our lives, evoking emotions that range from excitement to introspection. Unraveling the sentiments embedded within movie reviews can provide valuable insights into audience perceptions. This project delves into the intriguing world of sentiment analysis, specifically focusing on IMDB movie reviews. By employing advanced Natural Language Processing (NLP) techniques, we aim to decipher the sentiments hidden within textual reviews, ultimately discerning whether the feedback leans towards positivity or negativity. Through the exploration of both Multi-Layer Perceptron (MCP) models using Bag of Words and Convolutional Neural Networks (CNN) utilizing Embedding, we embark on a journey to extract and quantify the nuances of emotions that cinematic experiences evoke.

DATA

For this project, we leverage a dataset containing a curated collection of IMDB movie reviews, encompassing a diverse range of cinematic experiences. Sourced from the Kaggle platform, this dataset consists of 39723 movie reviews, I will use the first 10,000 movie reviews, each annotated with sentiment labels. These labels categorize the reviews into positive or negative sentiments, reflecting the audience's emotional response to the films. By harnessing this rich dataset, we are equipped to explore the intricacies of sentiment analysis and discern the sentiments embedded within the textual expressions of moviegoers. This dataset not only serves as the foundation for our analysis but also provides the key insights necessary to train, validate, and evaluate our models accurately.

IMPLEMENTATION

Our implementation involves deploying Convolutional Neural Network (CNN) and Multi-Layer Perceptron (MCP) models for sentiment analysis on IMDB movie reviews. Utilizing Bag of Words for MCP and Embedding techniques for CNN, we employ Python's TensorFlow framework for model creation, training, and evaluation. The project showcases proficiency in Natural Language Processing (NLP) concepts, emphasizing model selection, feature representation, and performance comparison.

dense_input	input:	[(None, 38007)]
InputLayer	output:	[(None, 38007)]

dense	input:	(None, 38007)
Dense	output:	(None, 50)

dense_1	input:	(None, 50)
Dense	output:	(None, 1)

MCP Model – A summary of inputs and outputs using neural bag-of-words model

CNN Model - A summary of inputs and outputs using single channel convolutional neural network

embedding_input	input:	[(None, 1480)]
InputLayer	output:	[(None, 1480)]

embedding	input:	(None, 1480)
Embedding	output:	(None, 1480, 100)

conv1d	input:	(None, 1480, 100)
Conv1D	output:	(None, 1473, 32)

max_pooling1d	input:	(None, 1473, 32)
MaxPooling1D	output:	(None, 736, 32)

flatten	input:	(None, 736, 32)
Flatten	output:	(None, 23552)

dense	input:	(None, 23552)
Dense	output:	(None, 10)

dense_1	input:	(None, 10)
Dense	output:	(None, 1)

SET

SCHOOL OF ENGINEERING
AND TECHNOLOGY

RESULT

	MCP	CNN
Training Accuracy	100.00%	100.00%
Test Accuracy	87.20%	87.40%

CONCLUSION

In this project focused on sentiment analysis of IMDB movie reviews, we utilized a dataset of 10,000 reviews to compare the performance of two distinct models: Multi-Layer Perceptron (MCP) and Convolutional Neural Network (CNN). Surprisingly, both models achieved remarkable training accuracies of 100%, indicating their capacity to learn from the training data. However, when tested on the held-out data, the MCP model attained an accuracy of 87.20%, while the CNN model exhibited slightly improved performance with an accuracy of 87.40%. These results underscore the effectiveness of both models in capturing sentiment patterns in movie reviews. The CNN model's marginally higher accuracy suggests its ability to leverage local features through convolutional layers, providing a nuanced understanding of sentiment nuances within text.

REFERENCES

- ChatGPT. <https://chat.openai.com/>
- Yasserh. (2021). IMDB Movie Ratings Sentiment Analysis. Kaggle. <https://www.kaggle.com/datasets/yasserh/imdb-movie-ratings-sentiment-analysis>