Movie reviews sentiment classification with semantic search

Abdulrahim Ahmadov
Tomáš Timko
Maroš Matej

Introduction

•**Objective:** Our research focuses on classifying movie reviews into three categories: Positive, Negative, and Neutral.
•**Dataset:** Rotten Tomatoes Movie Reviews dataset containing 1.44 million rows.

Methodology

Cleaned data by removing missing values and mapping sentiment labels ("fresh" to 0, "rotten" to 1).
Processed the large dataset in chunks of 10,000 rows for memory efficiency
Tokenized reviews with BERT tokenizer and truncated to 64 tokens.
Split data into Training (80%), Validation (10%), and Testing (10%) sets.
Fine-tuned bert-base-uncased using Adam optimizer (learning rate:2e-5) for 3 epochs
Evaluated performance with metrics: Accuracy, Precision, Recall, and F1-Score

Results

Epoch	Loss	Accuracy	Precision	Recall	F1-Score
1	0.23	0.90	0.90	0.90	0.90
2	0.22	0.91	0.91	0.91	0.91
3	0.25	0.91	0.91	0.91	0.91

We developed a robust sentiment classification model using a fine-tuned BERT architecture, achieving over 91% accuracy on movie review sentiment analysis. Additionally, our system supports semantic search, enabling contextual and nuanced retrieval of movie reviews based on meaning rather than just keywords.

Download the digital copy of this poster!

KKUI
Katedra kybernetiky a umelej inteligencie

Visuals