

An Automatic Movie Summary Generator

Luka Pavićević, Andrija Stanišić, Stefanela Stevanović

Abstract

Advisor: Slavko Žitnik

Introduction

In today's world, there is an abundance of films being produced every year. It can be difficult for audiences to keep up with all the latest releases. Movie fans often resort to reading summaries or synopses to get a sense of what a particular film is about before deciding whether or not to watch it. However, it can be a time-consuming and difficult task to manually summarise a long film script.

Here Natural Language Processing comes in. Through the use of Natural Language Processing techniques, we can automate the task of summarising film scripts, saving both time and effort.

In this project, our aim is to create a system that can accurately summarise film scripts into concise and informative synopses through the use of Natural Language Processing. By doing so, we hope to provide film enthusiasts with a quick and efficient way to get a sense of a film's plot, characters and key events. Ultimately, this will help them in their decision-making process when choosing which films to watch.

Related works

One of the related works in text summarization is the work by Aleš Žagar and Marko Robnik-Šikonja [1], who explored the use of various summarization approaches, including neural models, to produce short summaries from larger texts. In their work, they addressed the problem of selecting the most appropriate summarization model for a given text, and proposed a solution that uses a neural metamodel to automate the selection process based on the input document's representation. This work presents an innovative approach to text summarization, particularly in addressing the issue of model selection, which is important in ensuring the quality of generated summaries.

The article [2] discusses the performance of generic text

summarization algorithms on films and documentaries, using extracts from news articles produced by reference models of extractive summarization. The study aims to evaluate the quality of automatic summaries produced for films and documentaries, using well-known behaviors of news articles as a reference. Six text-based summarization approaches were used: Maximal Marginal Relevance (MMR), LexRank, Latent Semantic Analysis (LSA), Support sets, Key Phrase-based Centrality (KP-Centrality), and TextRank. The study found that LexRank and Support Sets had the best performance on movies of all the mentioned approaches.

The paper [3] proposes an ensemble deep learning model for generating movie scripts, which means that multiple models are combined to improve the accuracy and robustness of the system. The model uses the Ensemble-based Movie Script Generation (EMCG), where the Optimized hybrid script generation process using ensemble learning is performed by Bi-LSTM, GPT3, and GPT Neo X models, where the parameters of deep learning algorithms are optimized using the AI-CMO algorithm.

References

- [1] A. Žagar and M. Robnik-Šikonja. Slovenian text summarization models. 2022.
- [2] Marta Aparício, Paulo Figueiredo, Francisco Raposo, David Martins de Matos, Ricardo Ribeiro, and Luís Marujo. Summarization of films and documentaries based on subtitles and scripts. *Pattern Recognition Letters*, 73:7–12, 2016.
- [3] R. Dharaniya, J. Indumathi, and V. Kaliraj. A design of movie script generation based on natural language processing by optimized ensemble deep learning with heuristic algorithm. *Data Knowledge Engineering*, page 102150, 2023.