Event Extraction From Movies

Raj Chhari
Btech VI semester, Information Technology
Indian Institute of Information Technology
Allahabad, India iit2020010@iiita.ac.in

Abstract—Movies are a rich source of data for understanding human behaviour, storytelling, and cultural analysis. Extracting events from movies can provide valuable insights for various applications, such as scene summarization, script analysis, and movie recommendation systems. In this project, I propose a novel approach for event extraction from movies using natural language processing (NLP) techniques such as Named Entity Recognition(NER), Semantic Role Labeling (SRL), Clustering etc.

Index Terms—Event extraction, natural language processing (NLP), Named Entity Recognition (NER), Semantic Role Labeling (SRL), clustering, scene summarization, script analysis, movie recommendation systems.

I. Introduction

Event extraction is a natural language processing (NLP) technique that involves identifying and extracting relevant information about events or activities from unstructured text data such as news articles, social media posts, or emails. It is an important task in information extraction and has a wide range of applications, including event tracking, sentiment analysis, and recommendation systems. The goal of event extraction is to automatically identify and extract the key components of an event, such as the event type, the participants involved, the time and location of the event, and any other relevant information. This is typically done using machine learning algorithms that are trained on large amounts of annotated data.

Understanding events is important to understanding a narrative. Event complexity varies from one story to another and the ability to extract them is essential for multiple applications such as scene summarization, script analysis, and movie recommendation systems.

In this project I will be extracting all the events from the movies and with these particular events I will also show the sentence in which this event is happening. For this I will be using publically available “Movie Scripts Corpus” Dataset which contains almost 3000 movie script texts and annotations by structural elements, but I will be using less movies data for my project.

II. Literature Review

A. ”Neural Events Extraction from Movie Descriptions” byAlex Tozzo, Dejan Jovanovic, and Mohamed R. Amer:

This paper proposes a neural network-based approach for event extraction from movie descriptions. The paper addresses

Identify applicable funding agency here. If none, delete this.

the challenge of identifying events from textual descriptions of movies, which is useful for tasks such as movie summarization and recommendation systems. The proposed approach consists of a neural network model that uses a combination of convolutional and recurrent layers to extract event sequences from movie descriptions. The authors use a dataset of 48 movies and evaluate their approach using precision, recall, and F1score metrics. The results show that the proposed approach outperforms several baseline methods, including rule-based and supervised learning-based approaches. The paper also provides a detailed analysis of the performance of the proposed approach, showing that it is effective at identifying events from movie descriptions across different genres and languages. The authors also compare their approach to existing methods for event extraction from movie scripts and show that their approach achieves comparable performance.

B. ”Movie Scene Event Extraction with Graph Attention Net-work Based on Argument Correlation Information” by Qian Yi, Guixuan Zhang, Jie Liu and Shuwu Zhang:

This paper proposes a novel approach for event extraction from movies using a graph attention network based on argument correlation information. The paper addresses the challenge of identifying events in movie scenes, which is essential for tasks such as scene summarization and script analysis. The proposed approach consists of a two-stage process. In the first stage, the authors extract argument information from movie scripts using dependency parsing and semantic role labeling. They then construct a graph

representation of each scene, where nodes represent the arguments and edges represent the argument correlations. In the second stage, the authors use a graph attention network to extract events from the graph representation. The graph attention network is designed to focus on the most relevant arguments and correlations for each event, based on their semantic and syntactic properties. The authors evaluate their approach on a dataset of 40 Hollywood movies and compare their results with several baseline methods. The results show that the proposed approach achieves state-of-theart performance in terms of precision, recall, and F1-score metrics. The authors also conduct a qualitative analysis of the extracted events and show that their approach is able to capture a wide range of events, including physical actions, verbal communication, and emotional expressions.

C. "Lights, Camera, Action: Knowledge Extraction fromMovie Scripts" by Niket Tandon, Gerard de Melo, Abir De, and Gerhard Weikum:

This paper proposes a knowledge extraction approach for analyzing movie scripts. The paper addresses the challenge of extracting structured knowledge from unstructured textual data in the form of movie scripts. The proposed approach consists of a pipeline of several NLP techniques, including named entity recognition, semantic role labeling, and relation extraction. The authors use a dataset of 164 Hollywood movies and evaluate their approach using precision, recall, and F1score metrics. The results show that the proposed approach is effective at extracting a wide range of knowledge from movie scripts, including character interactions, locations, and events. The paper also provides a detailed analysis of the extracted knowledge, showing that it can be used to build a knowledge graph of the movie domain. The authors demonstrate several applications of the knowledge graph, including movie recommendation and script analysis

.

D. "Movie Script Summarization as Graph-based Scene Ex-traction" by Pavel G. Gorinski and Mirella Lapata:

This paper proposes a graph-based approach for movie script summarization. The paper addresses the challenge of summarizing movie scripts into a concise summary, which is essential for tasks such as movie recommendation and script analysis. The proposed approach consists of two main components. First, the authors use a dependency parser and named entity recognition to extract scenes from the movie script. Each scene is represented as a graph, where nodes represent entities and edges represent relations between entities. Second, the authors use a graph-based clustering algorithm to group similar scenes together and generate a summary of the movie. The authors evaluate their approach on a dataset of 200 movie scripts and compare their results with several baseline methods. The results show that the proposed approach is effective at summarizing movie scripts and outperforms the baseline methods in terms of several evaluation metrics. The paper also provides a detailed analysis of the extracted scenes and the resulting summary, showing that the graph based approach is able to capture important aspects of the movie plot and characters. The authors demonstrate several applications of the summarized movie, including movie recommendation and script analysis

.

E. "Extracting Character Information from Movie Script" bySoyoung Yoon, Sangwon Lee, Minyeop Choi, Shin Dong Hwan, and Kyumin Park:

This paper that proposes a method for character information extraction from movie scripts. The paper addresses the challenge of identifying and extracting character information, including character names, gender, and occupation, from unstructured movie script text. The proposed approach consists of a pipeline of several NLP techniques, including named entity recognition, part of-speech tagging, and co-reference resolution. The authors use a dataset of 32 movie scripts and evaluate their approach using precision, recall, and F1score metrics. The results show that the proposed approach is effective at extracting character information from movie scripts, with an F1-score of 0.87. The authors also analyze the extracted character information and demonstrate its usefulness for movie recommendation and script analysis. For example, the extracted character information can be used to build a knowledge graph of the movie domain, which can then be used to generate movie recommendations based on user preferences.

III. Methodology Here is

the proposed methodology:

1. The first step in my approach is text preprocessing,which involves cleaning and tokenizing the movie script or subtitles. Then I can use syntactic parsing to analyze the grammatical structure of the text and identify the syntactic roles of words and phrases. This is followed by named entity recognition, where I identify and classify entities such as characters, locations, and objects in the movie scenes

.

2. Next, I will employ semantic role labeling to identifythe semantic roles of words and phrases in relation to events. This allows me to identify the roles of entities as participants, such as the subject or object of an action, or as attributes, such as the time or location of an event. I will also use contextual information, such as co-occurrence patterns and verb semantics, to disambiguate between different senses of words and improve the accuracy of event extraction

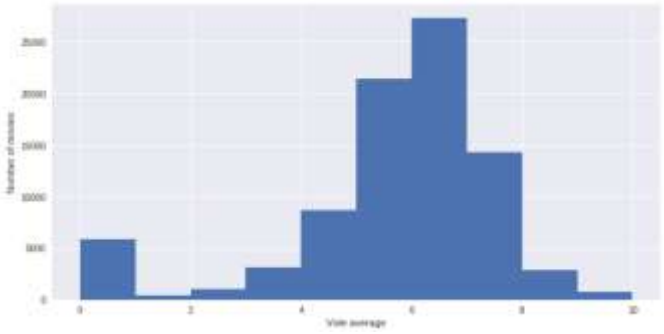
.

3. Finally, I aggregate the extracted events to create a comprehensive event timeline for the movie, capturing the

temporal order and duration of events. I will also apply event clustering techniques to group similar events together and identify patterns or themes in the movie scenes

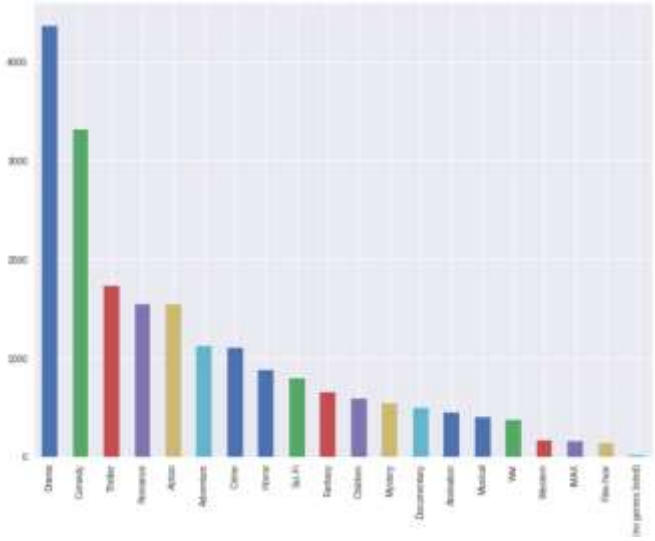
.

4. The extracted events can be further utilized for variousdownstream applications, such as scene summarization, script analysis.



IV. Result

	title	year	vote_count	vote_average	popularity	genres	ur
20896	Inception	2010	14075	8	29.108149	[Action, Thriller, Science Fiction, Mystery, A...	7.708258
20601	Inception	2010	14075	8	29.108149	[Action, Thriller, Science Fiction, Mystery, A...	7.708258
24256	The Dark Knight	2008	12269	8	123.167259	[Drama, Action, Crime, Thriller]	7.872005
15867	The Dark Knight	2008	12269	8	123.167259	[Drama, Action, Crime, Thriller]	7.872005
45602	Interstellar	2014	11187	8	32.213481	[Adventure, Drama, Science Fiction]	7.845814
36825	Interstellar	2014	11187	8	32.213481	[Adventure, Drama, Science Fiction]	7.845814
8424	Fight Club	1999	9678	8	63.869599	[Drama]	7.801264
8750	The Lord of the Rings: The Fellowship of the Ring	2001	8892	8	32.070725	[Adventure, Fantasy, Action]	7.573273
292	Pulp Fiction	1994	8670	8	140.950238	[Thriller, Crime]	7.904848
314	The Shawshank Redemption	1994	8358	8	51.645403	[Drama, Crime]	7.851923
13462	The Lord of the Rings: The Return of the King	2003	8226	8	29.324358	[Adventure, Fantasy, Action]	7.540308
4453	The Lord of the Rings: The Return of the King	2003	8226	8	29.324358	[Adventure, Fantasy, Action]	7.540308
351	Forrest Gump	1994	8147	8	49.307194	[Comedy, Drama, Romance]	7.542901
2813	The Lord of the Rings: The Two Towers	2002	7641	8	29.423537	[Adventure, Fantasy, Action]	7.519845
10990	The Lord of the Rings: The Two Towers	2002	7641	8	29.423537	[Adventure, Fantasy, Action]	7.519845



Acknowledgment

I would like to thank prof. Muneendra Ojha sir and all the Teaching Assistants who has provided me with valuable guidance, support, and encouragement throughout the project. Their expertise and insights have been invaluable in shaping the direction of this work. Finally, I would like to express my appreciation to the wider academic community for their contributions to the field of NLP, which have been a source of inspiration and knowledge for this project.

References

[1] Neural Events Extraction from Movie Descriptions" by Alex Tozzo, Dejan Jovanovic, and Mohamed R. Amer.

[2] https://aclanthology.org/W18-1507/

[3] Movie Scene Event Extraction with Graph Attention Network Based on Argument Correlation Information" by Qian Yi, Guixuan Zhang, Jie Liu and Shuwu Zhang

[4] https://www.mdpi.com/1424-8220/23/4/2285

[5] "Extracting Character Information from Movie Script" by Soyoung Yoon, Sangwon Lee, Minyeop Choi, Shin Dong Hwan, and Kyumin Park

[6] Lights, Camera, Action: Knowledge Extraction from Movie Scripts" by Niket Tandon, Gerard de Melo, Abir De, and Gerhard Weikum.