Movie Recommender

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Anisha Aggarwal, Shruti Jain, Ananya Menon, Ajay Sumanth, and Richard Wang

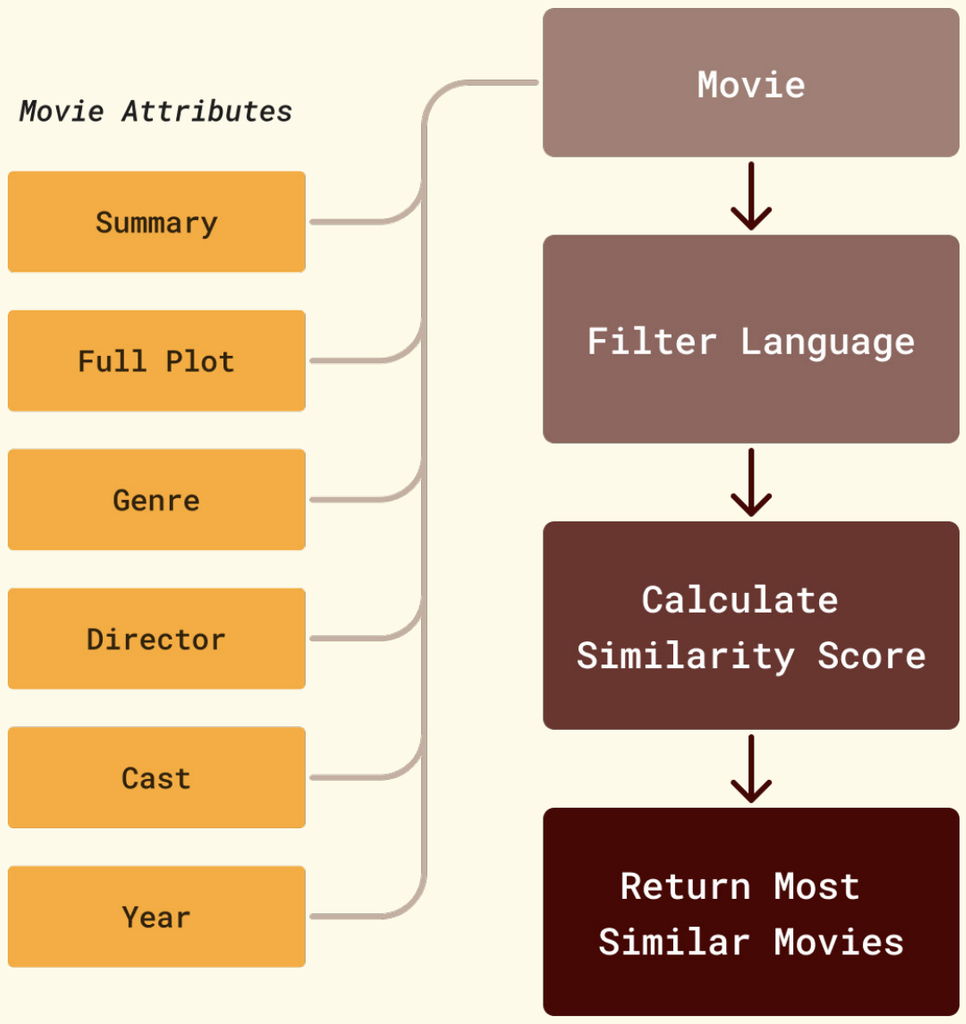
4

Problem Statement

- Users desire to consume media similar to what they have just watched
- Existing recommendations often fail to capture the essence of the original movie's plot and tend to be only loosely related
- There is no way for the user to choose which factors go into their recommendations

Our Solution

- A movie recommender that takes in a movie as the query and returns the top 25 similar movies
- Allows the user to change how important different features (Genre, plot, director, etc.) are



Methodology

- **Overall Algorithm**: Uses criteria like genre, plot, language, cast, director, and release year to recommend movies. Attributes are stored in dictionaries, and GloVe embeddings are used for plots to compare movies. Movies are filtered by genre and language to streamline the process.
- **Weighting Scheme**: Applies weights primarily to plot summaries and full plots, with substantial importance also given to genre. Director, cast, and release year have smaller weights. The weighting system has been optimized for best performance.
- **Calculating Similarity Scores**: Fixed-length embeddings for plot summaries and full plots are compared using cosine similarity. Additional scores are calculated for cast and director overlap, and release year differences are penalized. Genre similarity is assessed with Jaccard similarity.
- **User Adjusted Weights**: Users are able to adjust weights to refine recommendations, offering five levels of importance for each criterion.

Evaluation

Genre	Number of Movies	Average Precision@5	Average Precision@10
Action	6	0.8	0.72
Adventure	1	0.6	0.5
Animated	4	0.9	0.73
Comedy	6	0.83	0.73
Drama	15	0.8	0.67
Fantasy	3	0.6	0.53
Horror	1	1	0.9
Musical	1	0.6	0.4
Mystery	1	0.8	0.8
Romantic Comedy	1	1	1
Science Fiction	7	0.8	0.64
Superhero	2	0.9	0.75
War	2	0.6	0.55

Avg Top 5	Avg Top 10	
0.80	0.678	<-- Across 50 Movies

Future Work

- **Fine Tuning Weights:** Create different versions with different default weighting schemes and do A/B testing to fine tune the weights
- **Other Media:** Add other forms of media to recommend (books, TV shows, music)
- **Explanation:** Explain to the user the exact reason for why a recommendation was given
- **Keyword Search:** Filter by user inputted keywords

Dataset

- **Data Collection Method:** Gathered 35,000 Wikipedia movie summaries and metadata using a HuggingFace dataset, enabling genre-based filtering and similarity evaluations.
- **Data Annotation and Preprocessing Method**: Movies are categorized by genre, language, and other attributes. Plot summaries and full plots are cleaned and tokenized for analysis. Cast and director information is systematically organized.
- **Interesting Data Samples**: The dataset includes films of varied popularity and critical acclaim, multi-genre films, and foreign films in approximately 20 languages, providing a basis for evaluating and enhancing the recommender system's effectiveness.

Conclusion

- The recommendation system is designed to suggest movies based on objective characteristics such as plot, genre, release year, directors, cast, and language, without considering individual user preferences.
- A unique weighting scheme was developed to calculate similarity scores between movies, which has generally resulted in effective recommendations across most genres.
- The system struggles to provide accurate recommendations for drama films compared to other genres like comedy, horror, and superhero.
- The evaluation of the recommendation system's effectiveness faced challenges due to user bias, making it difficult to obtain an objective assessment of the quality of movie recommendations