

Movie Recommendation System Using NCF

Recommender Systems course project

Done by: Khalid Nimri (2140145), Aseel Suhail (2140197)

Abstract

This article introduces an innovative architecture for a movie recommendation system of rules that incorporates Neural Collaborative Filtering (NCF).

NCF leverages deep learning methods to accurately predict exploiter preferences based on their flic viewing patterns.

The system's multi-layered neural networks process user stimulus information and place hidden relationships, enhancing recommendation accuracy.

This approach addresses data thinness and continually integrates newly users, leading to precise and reliable motion-picture show recommendations.

Introduction

The age of digital streaming and a wide variety of preprogrammed content meant the recommendation systems become an integral part of navigating an overwhelming audience through movies, TV shows and other types of entertainment to display the material most timely to each individual. This project aims to build a feature-rich movie recommendation system learning by utilizing cutting-edge machine learning technologies to present customized recommendations for the users based on their preference and their activity.

Background/Related Work

-Content-Based Filtering: This method recommends items similar to those the user has liked in the past, based on item features like genres or actors. However, it often struggles to suggest new or diverse items outside the user's previous preferences.

-Collaborative Filtering: Divided into user-based and item-based approaches, this method recommends items based on the preferences of similar users or similarities between items. Despite its popularity, collaborative filtering faces issues like data sparsity and cold starts, where there is insufficient data for new

-Neural Collaborative Filtering (NCF): Neural Collaborative Filtering (NCF) represents a significant advancement in recommendation systems by leveraging deep learning to model complex user-item interactions. NCF combines:

1-Generalized Matrix Factorization (GMF): Captures linear user-item interactions.

2-Multi-Layer Perceptron (MLP): Models non-linear interactions through multiple neural network layers.

3-Fusion Layer: Merges the linear and non-linear representations to produce the final prediction.

The advantages of NCF include its flexibility in modeling interactions, scalability to large datasets, and enhanced personalization capabilities compared to traditional methods.

Approach

Data Collection and Preprocessing:

A giant data set that comprise the elements of movie names, genres, and summaries will be recorded together with the mood score of the particular movie.

The data cleaning is an important part as it involves replacement of anomalies and analysis of missing values with accurate data before the main project starts.

Model Training:

-Word2Vec: Put forward a task to a model which has the capacity to learn the same word vectors for movies as well as movie descriptions to have the word vector transfer from one movie to another to represent the semantic relations of the words in the movies

-TF-IDF Vectorizer: Apply TF-IDF transformation on movie descriptions for the purpose of understanding the words of vector labeling to make them relevant, for instance, key words for any similarity computing or matching process

-Neural Collaborative Filtering (NCF): We will be able to fully describe the interweaving relations between GMB and MLP networks in order to determine more precisely how linear and non-linear features of Item-User are influenced. Attempt to make use of data to the max by forcing the NN wings to integrate with the basic rating model However, the lack of funding can have a detrimental impact on entrepreneurial aspirations

Recommendation Generation:

Content-Based Recommendations: It is appropriate to get the cosine similarity based on the tf-idf vectors in order to identify the nearest neighbor movies to given reference movie by using the top facts and then advice the best of top movies.

Top-Rated Movies: Guide and imprint recommended films that belong to a certain genre and have high viewers' ratings, at least 4 out of 5 stars, when there are no films matching or related to viewers' selection.

User Interaction:

Concentrate on understanding the needs of the users and play the genre that they would like to hear out. Furthermore, it is a good idea to request them to fill in the average rate on that genre and then list up names of other movies. They can also list as many movies that they have loved previously.

A motion going through the carefully described film structure such as "title" providing "genre" as well, the "duration", "year" and "rating" , and a short description, will have to be shown dynamically.

Experiments:

- **IMDB Dataset:** Contains information on over 100,000 movies and TV shows, including titles, genres, descriptions, and user ratings.

Configuration Details:

- **Word2Vec:** Vector size = 100, window size = 5, min count = 1.

- **TF-IDF:** Standard configuration for text vectorization.

- NCF Model:

Embedding size: 50

Layers: Dense(128, ReLU), Dense(64, ReLU)

Optimizer: Adam

Loss function: Mean Squared Error

- Training Parameters:

Learning rate: Adjusted per model

Epochs: 10

Batch size: 64

Early stopping: Used to prevent overfitting

- Evaluation Metrics

Accuracy: Measures the correctness of recommendations

Precision: Fraction of relevant recommendations among the retrieved items.

Recall: Fraction of relevant items that were recommended.

Mean Squared Error (MSE): Evaluates prediction accuracy for ratings.

Results:

| Model configuration | Accuracy | Precision | Recall | MSE |
|---------------------|----------|-----------|--------|-------------|
| Word2vec + NCF | 0.85 | 0.87 | 0.82 | 0.15 |
| TF-IDF + NCF | 0.83 | 0.85 | 0.80 | 0.18 |

Discussion and Conclusion

Utilizing Word2Vec algorithm as well as TF-IDF, and Neural Collaborative Filtering (NCF) pieces of this puzzle, our movie recommendation system helps improve the personalized movie suggestions. Compared with Word2Vec + NCF, others also have advantages but this construction has a larger print, higher accuracy and precision. It is an intuitive hence, machine learning models of advanced skill are implemented so as to enhance the quality of recommendations and increase the satisfaction of the users.

References:

-Journal: <https://dl.acm.org/doi/10.1145/3038912.3052569>

-Dataset: <https://www.kaggle.com/datasets/kurtnakasato/imdb-100000-moviestvshows>

Team Contribution:

| | | |
|---------------------|----------------|--|
| Khalid Nimri | 2140145 | <ul style="list-style-type: none">• get_similar_movies function• recommend_movies function• ask_user function• Presentation slides• Some work on the final write up |
| Aseel Suhail | 2140197 | <ul style="list-style-type: none">• Train_word2vec function• NCF_model function• Most of the final write up. |