EE 541 – Computational Introduction to Deep Learning

Project Proposal

Yichen Wang, Rui Zhao

November 15, 2023

Project Title: Movies Recommendation System

Topic summary

Our objective is to develop a movie recommendation system in order to enhance the user's experience by providing movie suggestions. The core functionally of this system is that when user provides us with one specific movie they watched before, we are able to process the movie metadata such as genre, topics, languages, etc. Then we output a result of a comprehensive recommendation list that offers the user a selection of films closely related with their input, such that giving the user a smoother and more tailored viewing experience.

Introduction

- **Explain the significance of the problem. Who benefits from a solution?**
  Movie recommendation systems significantly enhance user experiences by providing personalized movie suggestions and saving time for searching. The system benefits consumers by simplifying movie selection and introducing them to new contents which may appeal to their interest.

- **Describe the impact on industry or to academic research. List examples, *e.g.*, Company X can benefit from using deep learning to do Y.**
  Such system offers substantial advantages to streaming platforms and movie distributors, increasing user engagement, possibly revenue. Additionally, they provide valuable insights for advertisers and content creators within the entertainment industry. For example, video streaming platforms like Amazon Prime and Netflix can benefit from deep learning methods to provide better user experience.

- **Provide (short) inclusions of any theoretical, applied, or algorithmic background necessary to understand this proposal.**
  We have several methods in mind for this movie recommendation system.

1. **Collaborative Filtering with Deep Learning:**
   Basically, this method relies more on the users. Let's say if we have two users who have watched the same movie, if one of the users watched a new movie and liked it, it is of high possibility that the other user who have a similar past viewing experience would also agree on this movie.

2. **Content-Based Filtering with Machine Learning:**
   Natural Language Processing (NLP): Techniques like TF-IDF or word embeddings for textual data analysis.
   Convolutional Neural Networks (CNNs): For analyzing visual content in movie posters.

3. **Feature Engineering and Dimensionality Reduction:**
   Principal Component Analysis (PCA): For reducing feature space dimensionality.

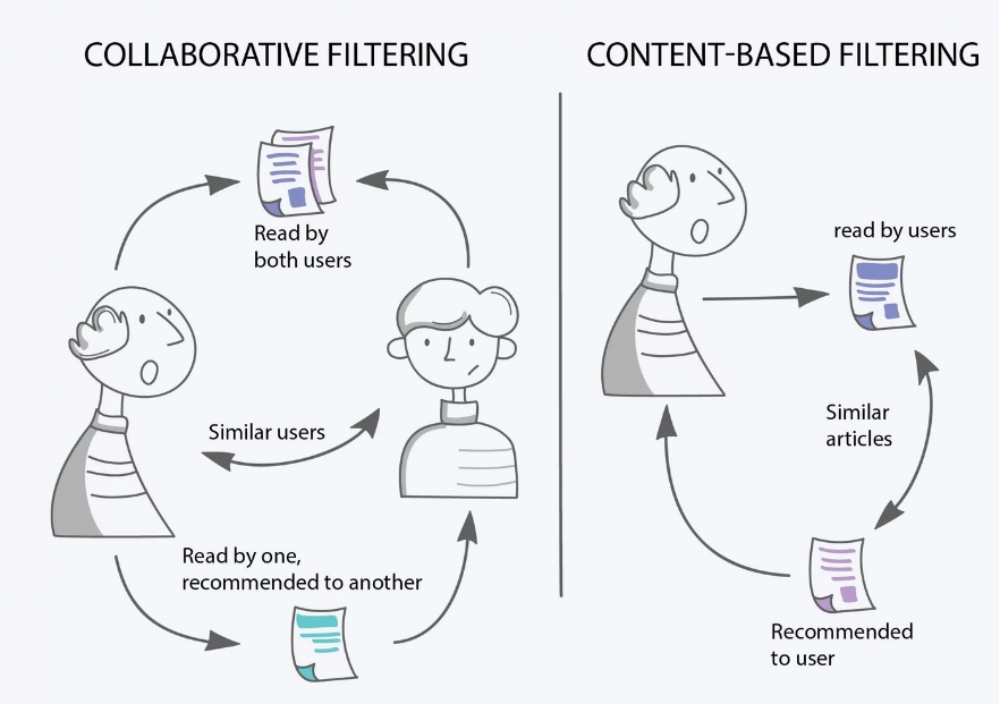

Figure 1: Demonstrates the concepts of Collaborative Filtering and Content-Based Filtering.

copyright: www.kdnuggets.com

Related Work

We found this paper doing a research on the movie recommendation system based on a hybrid approach by combining Content based filtering and Collaborative filtering, using Support Vector Machine as a classifier and genetic algorithm. (See reference No.2.)

Dataset description

https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset/data

These files contain metadata for all 45,000 movies listed in the Full MovieLens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote counts and vote averages.

This dataset also has files containing 26 million ratings from 270,000 users for all 45,000 movies. Ratings are on a scale of 1-5 and have been obtained from the official GroupLens website.

Architecture Investigation Plan

1. Perform data preprocessing and analyze all given data files and consider the features. Perform word embeddings if necessary.
2. Evaluate the necessity for feature engineering, such as normalizing or standardizing the data, and consider dimensionality reduction techniques like PCA. For movies lacking certain features, decide whether to discard them or assign a fixed value.
3. Divide the dataset into training, validation, and testing sets to ensure proper evaluation of the model's performance.
4. Experiment with Collabrative Filtering and Content-based Filtering. Compare their performance to determine the most effective approach for your recommendation system.

Estimated Compute Needs

Directly use Cuda 12.1.112 from Local Environment. Or use 11th Gen Intel(R) Core(TM) i9-11900H CPU.

Primary References and Codebase

We propose to build on the approach used in

- https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset/data
- S. Agrawal and P. Jain, "An improved approach for movie recommendation system," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2017, pp. 336-342, doi: 10.1109/I-SMAC.2017.8058367.