AI for Book Analysis

SABUDH

Literature Review & Progress

Presented by:

Jaykishan Padia

Shreya Chatterjee

Rohan Khosla

Thaidu Sai Vamshi

Souptik Dutta

Avik Agarwala

Project Mentor(s):

Jaybrata Chakraborty

Apoorav Mittal

Aashish Rana

Introduction

Our project aims to develop a sophisticated Book Recommendation System that combines content-based and collaborative filtering with a genre classifier, leveraging **machine learning and natural language processing** to deliver highly personalized and engaging book recommendations.

- **Addressing the Book Discovery Challenge**: Our project tackles the difficulty readers face in finding the right book amidst an overwhelming array of choices by developing an advanced Book Recommendation System.

- **Advanced Algorithm Integration**: The system utilizes a *combination* of collaborative filtering and content-based algorithms to provide sophisticated, personalized book recommendations.

- **Enhancing the Reading Experience**: The ultimate goal is to improve the reading experience by offering precise, engaging recommendations, simplifying book discovery, and fostering a deeper appreciation for reading.

Problem Statement

- Traditional book recommendation methods often provide limited accuracy and fail to fully analyze user preferences and book content.

- Lack of scalable and main-stream Book Recommendation Systems that can be used on a daily basis, and trusted on accuracy and effectiveness.

- If we can have personalised suggestions on Netflix and Youtube for movie buffs, why not something similar for all the book worms out there?

The goal is thus to create a hybrid book recommendation system which provides optimal recommendations personalized for each user.

Project objectives

1

Develop a personalized book recommendation platform using Natural Language Processing and Deep Learning techniques.

2

Integrate content-based filtering, collaborative filtering, and a genre classifier to provide tailored book recommendations to users.

3

Create a user-friendly interface for book discovery, rating, and feedback, with deployment on AWS for scalability and accessibility.

Literature review

- **Content-based Filtering**:

  1. Uses the attributes or features of an item to recommend other items similar to the user's preferences.

- **Collaborative Filtering**:

  1. Recommends items based on the information collected from multiple users.

  2. This approach uses similarity of user preference behavior: Given previous interactions between users and items, recommender algorithms learn to predict future interaction.

- **Hybrid recommender systems**:

  1. Combine multiple techniques to improve recommendation accuracy and overcome limitations of individual methods.

Literature review

Common Techniques

- **Cosine Similarity:** Cosine similarity is a metric, helpful in determining the similarity of data objects.

- **Weighted Hybridization:** Combines the scores from different recommendation techniques and assigns weights to generate a final recommendation.

- **Term Frequency-Inverse Document Frequency (TF-IDF):** Calculates how relevant is a word in a series or corpus is to a text.

- **K-Means Clustering:** Groups similar items or users together to make recommendations.

- **Artificial Neural Networks (ANNs):** Used for learning complex patterns in data to improve recommendation accuracy.

- **Neuro-Fuzzy Systems:** Combine neural networks and fuzzy logic to model complex relationships and handle uncertainty in recommendations.

Literature review

- **API**: Open Library API for data retrieval.

- **Count Vectorizer**: Used for converting data into vectors.

- **Bert Transformer**: Represent text as a sequence of vectors.

- **Matrix Factorization (e.g., Singular Value Decomposition - SVD)**: Reduces the dimensionality of the user-item interaction matrix to discover latent features that can predict user preferences more accurately.

- **Keras, Scikit-Learn, PyTorch**: For Machine Learning models.

Literature review

Results:

Pattern-based hybrid book recommendation system using semantic relationships Authors: Fikadu Wayesa, Mesfn Leranso, Girma Asefa & Abduljebar Kedir	**Precision:** 0.6384141 **Recall:** 0.404274
A systematic review and research perspective on recommender systems Authors: Deepjyoti Roy & Mala Dutta	**Agreement Rate (Neuro-Fuzzy Model):** 85.88%.
Personalized Book Recommendation System using Machine Learning Algorithm Authors: Dhiman Sarma, Tanni Mittra, Mohammad Shahadat Hossain	**F1 Score:** 52.84293

Literature review

Datasets

Open Source websites and Data Sources:

- Open Library (Data Dump: 12.4G)
- Goodreads (10,000 Books Dataset)
- Data was collected using a cross-platform data collection API from:
  1. Twitter (Dataset size: 12,098,339 tweet messages)
  2. MetaFilter dataset (size: 229,401 threads)
  3. Yahoo Answers dataset (size: 7,941,404 threads)
  4. Facebook data (Dataset size: 10 datasets each having 200 to 2000 Facebook posts)

Literature Review

- **Cold Start Problem in Collaborative Filtering**: Difficulty in recommending items to new users or suggesting new items due to a lack of initial data.

- **Divide by Zero Error**: The paper mentions an issue with a new collaborative filtering technique, where a divide by zero error occurs when users give the same ratings to different items, affecting the recommendation accuracy.

- **Scalability:** Handling large datasets efficiently is challenging as the number of users and items increases.

- **Sparsity:** Many users rate only a few items, leading to incomplete data, which affects the system's accuracy.

- **Synonymy and Latency Problems**: Issues arise when similar items are listed differently or when new items are not immediately available for recommendation.

Project Progress So Far

Completed Tasks

- Literature Review

Initial Findings

- Different techniques to create a recommendation system.

- Comparison of the effectiveness of various algorithms.

- Different data sources which can be used to gather data for training the system.

Next Steps

Immediate Goals

- Data Collection
- Data Preprocessing and cleaning

Long-Term Plans

- Create a hybrid of content-based and collaborative-based book recommendation system along with a GUI using Streamlit.

- Investigate the use of deep learning and other advanced methods for better feature representation and more accurate recommendations.

- Deployment on AWS for scalability and accessibility.

Thank you